

```

1 ATGAGCCAGC CCAGGCCCCG CTACGTGGTA GACAGAGCCG CATACTCCCT
51 TACCCTCTTC GACGATGAGT TTGAGAAGAA GGACCGGACA TACCCAGTGG
101 GAGAGAAACT TCGCAATGCC TTCAGATGTT CCTCAGCCAA GATCAAAGCT
151 GTGGTGTGTT GGCTGCTGCC TGTGCTCTCC TGGCTCCCCA AGTACAAGAT
201 TAAAGACTAC ATCATTCCCTG ACCTGCTCGG TGGACTCAGC GGGGGATCCA
251 TCCAGGTCCC ACAAGGCATG GCATTTGCTC TGCTGGCCAA CCTTCCTGCA
301 GTCAATGGCC TCTACTCCTC CTTCTTCCCC CTCCTGACCT ACTTCTTCCT
351 GGGGGGTGTT CACCAGATGG TGCCAGGTAC CTTTGCCGTT ATCAGCATCC
401 TGGTGGGTAA CATCTGTCTG CAGCTGGCCC CAGAGTCGAA ATTCCAGGTC
451 TTCAACAATG CCACCAATGA GAGCTATGTG GACACAGCAG CCATGGAGGC
501 TGAGAGGCTG CACGTGTCAG CTACGCTAGC CTGCCTCACC GCCATCATCC
551 AGATGGGTCT GGGCTTCATG CAGTTTGGCT TTGTGGCCAT CTACCTCTCC
601 GAGTCCTTCA TCCGGGGCTT CATGACGGCC GCCGGCTGC AGATCCTGAT
651 TTCGGTGCTC AAGTACATCT TCGGACTGAC CATCCCCTCC TACACAGGCC
701 CAGGGTCCAT CGTCTTTACC TTCATTGACA TTTGCAAAAA CCTCCCCAC
751 ACCAACATCG CCTCGCTCAT CTTGCTCTC ATCAGCGGTG CCTTCTGGT
801 GCTGGTGAAG GAGCTCAATG CTCGCTACAT GCACAAGATT CGCTTCCCCA
851 TCCCTACAGA GATGATTGTG GTGGTGGTGG CAACAGCTAT CTCCGGGGGC
901 TGTAAGATGC CCAAAAAGTA TCACATGCAG ATCGTGGGAG AAATCCAACG
951 CGGGTTCCCC ACCCCGGTGT CGCCTGTGGT CTCACAGTGG AAGGACATGA
1001 TAGGCACAGC CTTCTCCCTA GCCATCGTGA GCTACGTCAT CAACCTGGCT
1051 ATGGGCCGGA CCCTGGCCAA CAAGCACGGC TACGACGTGG ATTCGAACCA
1101 GGAGATGATC GCTCTCGGCT GCAGCAACTT CTTTGGCTCC TTCTTTAAAA
1151 TTCATGTCAT TTGCTGTGCG CTTTCTGTCA CTCTGGCTGT GGATGGAGCT
1201 GGAGGAAAAT CCCAGGTGGC CAGCCTGTGT GTGTCTCTGG TGGTGATGAT
1251 CACCATGCTG GTCCTGGGGA TCTATCTGTA TCCTCTCCCT AAGTCTGTGC
1301 TAGGAGCCCT GATCGCTGTC AATCTCAAGA ACTCCCTCAA GCAACTCACC
1351 GACCCCTACT ACCTGTGGAG GAAGAGCAAG CTGGACTGTT GCATCTGGGT
1401 AGTGAGCTTC CTCTCCTCCT TCTTCCTCAG CCTGCCCTAT GGTGTGGCAG
1451 TGGGTGTCTC CTTCTCCGTC CTGGTCGTGG TCTTCCAGAC TCAGTTTCGA
1501 AATGGCTATG CACTGGCCCA GGTTCATGGAC ACTGACATTT ATGTGAATCC
1551 CAAGACCTAT AATAGGGCCC AGGATATCCA GGGGATTAAA ATCATCACGT
1601 ACTGCTCCCC TCTCTACTTT GCCAACTCAG AGATCTTCAG GCAAAAGGTC
1651 ATCGCCAAGA CTGTCTCCCT GCAGGAGCTG CAGCAGGACT TTGAGAATGC
1701 GCCCCCACC GACCCCAACA ACAACCAGAC CCCGGCTAAC GGCACCAGCG
1751 TGTCCTATAT CACCTTCAGC CCTGACAGCT CCTCACCTGC CCAGAGTGAG
1801 CCACCAGCCT CCGCTGAGGC CCCCGGCGAG CCCAGTGACA TGCTGGCCAG
1851 CGTCCCACCC TTCGTACCT TCCACACCCT CATCCTGGAC ATGAGTGGAG
1901 TCAGCTTCGT GGAATTGATG GGCATCAAGG CCCTGGCCAA GCTGAGCTCC
1951 ACCTATGGGA AGATCGGCGT GAAGGTCTTC TTGGTGAACA TCCATGCCCA
2001 GGTGTACAAT GACATTAGCC ATGGAGGCGT CTTTGAGGAT GGGAGTCTAG
2051 AATGCAAGCA CGTCTTTCCC AGCATACATG ACGCAGTCCT CTTTGCCAG
2101 GCAAATGCTA GAGACGTGAC CCCAGGACAC AACTTCCAAG GGGCTCCAGG
2151 GGATGCTGAG CTCTCCTTGT ACCACTCAGA GGAGGACATT CGCAGCTACT
2201 GGGACTTAGA GCAGGAGATG TTCGGGAGCA TGTTTCACGC AGAGACCCTG
2251 ACCGCCCTGT GA

```

FEATURES:

Start Codon: 1

Stop Codon: 2260

HOMOLOGOUS PROTEINS:

Top BLAST Hits:

	Score	E
gb AAF81911.1 AF279265_1 (AF279265) putative anion transpor...	476	e-133
gb AAF71715.1 AF230376_1 (AF230376) prestin [Meriones ungui...	471	e-131
ref NP_000432.1 pendrin [Homo sapiens] >gi 11421915 ref XP...	451	e-125
ref NP_035997.1 Pendred syndrome homolog (human); Pendred'...	448	e-124
ref NP_062087.1 Pendred syndrome homolog (human) [Rattus n...	447	e-124
ref NP_067328.1 down-regulated in adenoma [Mus musculus] >...	434	e-120
ref NP_000102.1 down-regulated in adenoma protein [Homo sa...	418	e-115
sp O70531 DTD_RAT SULFATE TRANSPORTER (DIASTROPHIC DYSPLASI...	365	1e-99
ref NP_000103.1 sulfate anion transporter 1; Diastrophic d...	362	1e-98
ref NP_031911.1 diastrophic dysplasia [Mus musculus] >gi 2...	357	4e-97

BLAST to dbEST:

	Score	E
gi 8630793 /dataset=dbest /taxon=960...	523	e-146

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

Expression information from BLAST dbEST hits:

gi|8630793 Human head-neck

Expression information from PCR-based tissue screening panels:

Human fetal lung

```

1 MSQPRPRYVV DRAAYSLTLF DDEFEEKDRT YPVGEKLRNA FRCSSAKIKA
51 VVFGLLPVLS WLPKYKIKDY IIPDLLGGLS GGSIQVPQGM AFALLANLPA
101 VNGLYSSFFP LLTYFFLGGV HQMVPGTFAV ISILVGNICL QLAPESKFQV
151 FNNATNESYV DTAAMEAERL HVSATLACLT AIIQMGLGFM QFGFVAIYLS
201 ESFIRGFMTA AGLQILISVL KYIFGLTIPS YTGPGSIVFT FIDICKNLPH
251 TNIASLIFAL ISGAFLVLVK ELNARYMHKI RFPIPTMIV VVVATAISGG
301 CKMPKKYHMQ IVGEIQRGFP TPVSPVVSQW KDMIGTAFSL AIVSYVINLA
351 MGRTLANKHG YDVDNQEML ALGCSNFFGS FFKIHVICCA LSVTLAVDGA
401 GGKSQVASLC VSLVVMITML VLGIIYLYPLP KSVLGALIAV NLKNSLKQLT
451 DPYYLWRKSK LDCCIWVVSF LSSFFLSLPY GVAVGVAFSV LVVVFQTQFR
501 NGYALAQVMD LDIYVNPITY NRAQDIQGIK IITYCSPLYF ANSEIFRQKV
551 IAKTVSLQEL QQDFENAPPT DPNNNQTPAN GTSVSYITFS PDSSSPAQSE
601 PPASAEAPGE PSDMLASVPP FVTFTLILD MSGVSFVDLM GIKALAKLSS
651 TYGKIGVKVF LVNIHAQVYN DISHGGVFED GSLECKHVFP SIHDAVLFAQ
701 ANARDVTPGH NFQGAPGDAE LSLYDSEEDI RSYWDLEQEM FGSMFHAETL
751 TAL

```

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

Number of matches: 3

```

1 153-156 NATN
2 156-159 NESY
3 580-583 NGTS

```

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 2

```

1 45-47 SAK
2 445-447 SLK

```

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 11

```

1 18-21 TLFD
2 158-161 SYVD
3 240-243 TFID
4 365-368 SNQE
5 459-462 SKLD
6 556-559 SLQE
7 635-638 SFVD
8 691-694 SIHD
9 722-725 SLYD
10 726-729 SEED
11 732-735 SYWD

```

[4] PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site

Number of matches: 2

```

1 7-15 RYVVDRAAY
2 447-454 KQLTDPYY

```

[5] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 10

1	77-82	GGLSGG
2	78-83	GLSGGS
3	89-94	GMAFAL
4	103-108	GLYSSF
5	335-340	GTAFFSL
6	435-440	GALIAV
7	481-486	GVAVG
8	485-490	GVAFFSV
9	581-586	GTSVS
10	681-686	GSLECK

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	51	71	0.893	Putative
2	82	102	1.020	Certain
3	107	127	1.729	Certain
4	130	150	1.497	Certain
5	186	206	1.723	Certain
6	228	248	1.517	Certain
7	256	276	1.898	Certain
8	288	308	1.252	Certain
9	338	358	1.568	Certain
10	383	403	1.304	Certain
11	412	432	2.345	Certain
12	469	489	1.997	Certain
13	619	639	1.146	Certain

bioRxiv preprint doi: <https://doi.org/10.1101/008463>; this version posted September 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

BLAST Alignment to Top Hit:

>gb|AAF81911.1|AF279265_1 (AF279265) putative anion transporter 1 [Homo sapiens]

Length = 738

Score = 476 bits (1224), Expect = e-133

Identities = 263/724 (36%), Positives = 428/724 (58%), Gaps = 36/724 (4%)

Frame = +3

Query: 54 LFDDEFEEKDR--TYPVGEKLRNAFRCSAKIKAVVFGLLPVLSWLPKYKIKDYIIPDLL 227
L + EE R + P + R +CS A+ A++ LPVL WLP+Y ++D+++ DLL
Sbjct: 15 LNQEHLLEELGRWGSAPRTHQWRTWLQCSRARAYALLLQHLPVLVWLP RYPVRDWLLGDLL 74

Query: 228 GGLSGGSIQVPQGMFAFALLANLPAVNGLYSSFFPLLTYFFLGGVHQMVPGTFAVISILVG 407
GLS +Q+PQG+A+ALLA LP V GLYSSF+P+ YF G + GTFAV+S++VG
Sbjct: 75 SGLSVAIMQLPQGLAYALLAGLPPVFGLYSSFYPVFIYFLFGTSRHSISVGTFAVMSVMVG 134

Query: 408 NICLQLAPESKFQVFNNATNESYVDTAAMEAERLHVSATLACLTAIIQMGLGFMQFGFVA 587
++ LAP+ A N+S ++ A +A R+ V++TL+ L + Q+GLG + FGFV
Sbjct: 135 SVTESLAPQ-----ALNDSMINETARDAARVQVASTLSVLVGLFQVGLGLIHFGFVV 186

Query: 588 IYLSSEFIRGFMTAAGLQILISVLKYIFGLTIPSYTGPGSIVFTFIDICKNLPHTNIASL 767
YLSE +RG+ TAA +Q+ +S LKY+FGL + S++GP S+++T +++C LP + + ++
Sbjct: 187 TYLSEPLVRGYTTAAAVQVFSQ LKYVFG LHLSSHSGPLSLIYTVLEVCKWLPQSKVGTV 246

Query: 768 IFALISGAFLVLVKELNARYMHKIRFPIPTMIVVVVATAISGGCKMPKKYHMQIVGEIQ 947
+ A ++G LV+VK LN + ++ PIP E++ ++ AT IS G + ++ + +VG I
Sbjct: 247 VTAAVAGVVLVVVKLLNDK LQQQLPMPGELLTLIGATGISYGMGLKHRFEVDVVGNI 306

Query: 948 RGFTPTVPSPVVLQWKDMIGTAFSLAIVSYVINLAMGRTLANKHGYDVDSNQEMIALGCSN 1127
G PV+P + ++G+AF++A+V + I +++G+ A +HGY VDSNQE++ALG SN
Sbjct: 307 AGLVPPVPAPNTQLFSKLVGSAFTIAVVGFAIAISLGKIFALRHGYRVDSDNQELVALGLSN 366

Query: 1128 FFGSFFKIHVICCALSVTLAVDGAGGKSQVASLCVSLVVMITMLVLGIYLYPLPKSVLGA 1307
G F+ + C++S +L + GG SQVA SL +++ ++ LG + LPK+VL A
Sbjct: 367 LIGGIFQCFPVSCSMRSILVQESTGGNSQVAGAISSLFILLIIVKLGE LFDLPKAVLAA 426

Query: 1308 LIAVNLKNSLKQLTDPYYLWRKSKLDCCIWVVSFLSSFFLSLPYGVAVGVAFSVLVVVFQ 1487
+I VNLK L+QL+D LW+ ++ D IW+V+F ++ L+L G+ V V FS+L+VV +
Sbjct: 427 IIVVNLKGMRLQLSDMRSLWKANRADLLIWLVTFTATILLNLDLGLVVAVIFSLLL VVVR 486

Query: 1488 TQFRNGYALAQMVDTDIYVNPKTYNRAQDIQGIKIITYCSPLYFANSEIF----- 1637
TQ + L QV DTDIY + Y+ A++++G+K+ + +YFAN+E +
Sbjct: 487 TQMPHYSVLGQVPDTDIYRDVAEYSEAKEVRGVKVRSSATVYFANA EFYS DALKQRCGV 546

Query: 1638 -----RQKVIK--TVSLQELQQDFE-NAPPTDPNNNQTPAN-GTSVSYI----- 1760
++K++ K + L++LQ++ + P N TS+ +
Sbjct: 547 DVDFLISQKKLLKKQEQLKLKQLQKEEKLKQAASPKGASVSINVTSL EDMRSNNVED 606

Query: 1761 -----TFSPDSSSPAQSEPPASAEAPGEPDMLASVPPFVTFHTLILDMSGVSFVDLMGI 1925
S D A + ++AP + S + A P FH+LILD+ +SFVD + +
Sbjct: 607 CKMMQVSSGDKMEDATANGQEDSKAP-DGSTLKALGLPQPDFHSLILD LGALS FVDTVCL 665

Query: 1926 KALAKLSSTYGKIGVKVFLVNIHAQVYNDISHGGVFEDGSLECKHVFP SIHDAVLFAQAN 2105
K+L + + +I V+V++ H+ V + + G F D S+ KH+F S+HDAV FA +
Sbjct: 666 KSLKNIFHDFREIEVEVYMAACHSPVVSQLEAGHFF-DASITKKHLFASVHDAVT FALQH 724

Query: 2106 ARDV 2117
R V
Sbjct: 725 PRPV 728

FIGURE 2, page 3 of 4

[illegible]

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00916	1/1	187	497 ..	1	328 []	254.5	1.5e-72
PF00189	1/1	651	661 ..	79	89 .]	3.3	8

1 CTGGGTTTCCT ATGTGGGGAG GTCATGCTCC CCACTCATTG AGCCCCCCCCA
51 GGCAAACCAC CTGGACAGCC AGACCCATGC AGACTCTGGA GCAGGTGGAG
101 AGGAAGAGTG AGACCACCCC GCCTCACGGG CGGTGAAGGG CCGGCAGCCT
151 CTGAATAGTC TCTGCTAGGA GGTAGAAAGC ACCCTCCCAT CTTAATCATA
201 GTAATCATCG CCACTACCAT TTAATGGGTG CCTATAAAAG GCCAGCCTCT
251 TCATACACAT GATCTCACTG AATCCTCATA GCATCTGCCT GCGACTGTTA
301 TTATCCCAT TTACAGATGA AGAAACTGAA TCTTTGAACC CAGGTCATCT
351 GGCTCTCAAA CTGTGCTGT TTTCCCTAAG CCACCCGGTC TCTCATTTCT
401 CCCACTGAAA TGTCTCACAT GCCATTGCC TTAATCATTT CTGCCCATGT
451 CTCCTCCAAA ACACCATTTA TCAATTGCT TCAACAAGTAT GTGTGAGTA
501 CACACTAAGG GCCAGGCGAG GGGCTGGGCA CAGGCGCTGG GGGTAGGTTT
551 ATTCTCCAC CTTCGCTTCT GCTGGGTATC ACCTGTGGGG TCTTGCCGGG
601 CATCCACCC TCACCTGTAG TTCAAGTGA CCTTGGGATC CCAAGACCAA
651 ATGAATGGAA TGACCAGCC CAGCCTTCAC CAACTTGAGC ACAATCTTAT
701 TCATAATAGA AACTCACATT TGCATCACAC TTTACATTTT ACACAACCCC
751 TTCTTATCCA TTAATCATTT TGATCTTCAC AACAACCTTG TGAGATATGT
801 CTGTTACTCC CACTTTAGTG ATACAGAATC TGAGGTTTGA AAAGTAATGC
851 TGACCATTCT GCCTCATTAA TAAAAGCAGG ATTAACCCAG GCTCCTGGAC
901 CTTTCCACAA AAGGCATTAA GCAACCTGCT CCCCTCTGAC AACCTCCCTT
951 GTCACCAGG CTCTCCTCTG GGAAGTTGGG GGCATCTCTA GCCCCCAAGT
1001 AGTTACTCAT TTTCAACCCC ATCTCAAATC TTTTGCCAAA CTGGCCACAG
1051 CCACCCACCA CTCCCCACCT CCCAGATACA AATCCTCACT CTAAGCCTTC
1101 CCCATCTCTT TCTTCTCTGT CCTTCTTTCT CTGTGGTCTT CTGAGCAACT
1151 TCTCCAGCT CTGGGAGGTA GAGGGGAGGT GGGAGACCCA GTAATTGGAA
1201 GAGGGAGGGG GAAAGGTTCC TACAGGGAAC TCCTCCGGGC CTCAGGGGCC
1251 CTGGCACTCA GCTCTGCCCC TCTCAGCTCC TGGAACGTCA GCCAGGTTGC
1301 GCAAAAAGTG AGGAGGAGAG GAGCGGCAGT ACACAAGGGT GGGGGAAAGA
1351 TTAGGCACAG GAAGCCGTGG GAGAGAGAGC CGGCAGGTGG ACCATCCTGG
1401 TTTCCCCACA CACACCATTG TCCCCCTGGG AAACCTGTTG GTGAAGTTCT
1451 AGATGTCTTA TCCAAGAAGG GTCTCTTGA GGTCACTCTA GCTATCCCC
1501 TGCCTCTAGG CAAGCTGTTT TCTGTTTCTT CCAAGCTGAC TGGCTGAATG
1551 GTAGGAGCCT TTCTGCCAGG GAAACTAAGG TCTGGGAAGG GAGTATGGCT
1601 TGTGGGGACA CCAGGGGTCA GGGGAGGGGA GGGTCCACCT GCTGAATCAA
1651 GTGGGGCCTC CTGCCCTCGT GATTCCCTTT TGCTTGGTGC TCAGTGGGGG
1701 TGATGGTGAC GCCACAGGTG TGGAGTGCCA GCCACGTGCT GAGCGCCAAG
1751 CAAAACAGCC AGGGTGAGTG TATGCATCAT CAGTGCTGG GAAGGAAGGC
1801 CACTGCGAGC AGGGAGTCTG ACGGAAAAAC TTGACAGAGG GAAGGGAGGC
1851 ACCTTGCTTT ATCGGGGCGG GGAAGGCCAG AATAAACTC TGCTACTGCA
1901 AGGACCAGAG AGAGAAGGCC TGGGCTGGCA CTAGGGAGGG ATGTTCCCTC
1951 ACCCTCCCTT CCTCTGCTTC TCCCAAAGCT TGTAAATGCC CCAGATATGA
2001 GCCAGCCAG GCCCCGCTAC GTGGTAGACA GAGCCGCATA CTCCCTTACC
2051 CTCTTCGACG ATGAGTTTGA GAAGAAGGAC CGGACATACC CAGTGGGAGA
2101 GAAACTTCGC AATGCCCTCA GGTAAGTGGT CCAGAGCCCA GACTTCTGCC
2151 TCTCTGCTC CCTACCAAAA TCCTTTCTGC ACCAGGACAC GGCTTCTGCA
2201 CTGGTATCCC TAAGATGGGG TTAAGGGAAG CCCTGGGGAA GTGAGGTTCT
2251 GAATGATGAA TTTAAGATCC TACAACCTCA TCTGTACTGA GACCCCAAGG
2301 GAGGATGGGG AGCAGGAGCA AGAACCATCC AGAAGGGTTA TATGGCATTC
2351 CCAAACCCCT GCATGGCATC TCCCATATTC TCAATTCACC CGGGTCTCTC
2401 TGGGTTTGT T AAGGCATGGT AGATGAGCAT CTACGTTATG GAGGGGTGGG
2451 GAGCATCAGA GCCCTTACTC CATGCCCTGT TCCCTCCTTA CAAAAAATAC
2501 CTGAAGTTAC CATCACCCCA GGTTCCTTGT CCTTTCCCTC CCGGATGTTT
2551 CTCTCTCCAC TTGGTCCAGA GAATGCCAAA AGGAGGCCCT AAATTTCTGA
2601 ACTTTCCTGA GGGGACCTAC CAGGGTGTAG TCCTACCAGC GCCAGGGTCT
2651 TTTCCACTCT CATCTCCCTG GAAATGCGAT GGTGGGTATG AAACCTTGTC
2701 CCTAAGTAGG CGCTACACAA GGTGATCCAT ACCCACACCC CAGGAGGCTG
2751 GGGCTGCGGG TGTCACCCCT CCCATTCCCA GACTCCTGGC AGACCTCCTC
2801 TGGCCAGCT ATAGGCCAAC TCACTCTCCC TCACTCCCTT GGGGAAACGG
2851 CTGATTCACT TACCTGGATT GAGGTCACTG GCAATGGCTG AAGTGGAGAC
2901 GCAGGTGGAA CTGGTTTCAAG CCGGGGGAAT CACCCACTTG AGTTGTACT
2951 AAAAGCCCCA GCCCAGCCCT GTTTCTCTTG GGAGGCTCCA TTTCTGCCCA
3001 GTTACAGTCT GTCTCACAG CTGTGCTCCT CAGACAGGTG GTCTCTGCCA
3051 GTCTTTGTGC CCAAGACTTT AGGGCACAAA GTCTGAGGAT GAGAAGATCT
3101 GCTATTGTCC TAAAAGATTA GGATAATGAA AGCTGTAAAG GGATATAGCA

FIGURE 3, page 1 of 20

```

3151 AACTAACAAT TCCTATGATA CTGGCATGAG AGCCTTGAAC AGTGCCTGGC
3201 ATAGAGAAGG TGCACCAATA AATATTTGTT TCATGAATGA ATGAATGAAT
3251 GAATGTCTAG AAAGCTAATC CCTCTCAGCC TCTGTTTCCA GTTCTTCTTT
3301 CAAGCTTCAG ATTGCTTTGC CCAACATACA GCAGACTTGC AAGTAAGGTT
3351 GGGCATGGAC TAGCCCTCAA ATGAGTTGTT TTTCTTTCCC TAGCCAGCTC
3401 TCTATTCTATA AGTCCGGCTT TCTCTGCCAC AAACAGACCT GATGGAGCCC
3451 CTGCAGGGCT GGTCTCTCTCT TCAAGCAAGG CTTTAGAGTT GCATTAAGCA
3501 ATTTATCCCC CGTCCACCTC CCCTTCCAGC ATCCCAGGGA TGGCAGAGGC
3551 ACCCATGAGC CCCAGAAGGG ACAGGGGGTA AGATATTGAT GATGATGCTT
3601 TTTCTTGGAG TGTTAGTTGG AAGAGAAAAT CTGCCCAGAC TTTCCAAGGT
3651 ACAAAGCATT GTTTTGTGTT GTTTCAGTCT TGGGTGACAT CCAGGGGACC
3701 GAGTGTGAGG GAAACTATTG TTGAGCAAGA GCAAAGAGCA GGAATTGGTG
3751 CTGGGCAGGA AAGGAAGCCT CATCAGAGCA GGCCAGTGAG TCACCAAATG
3801 GGCCCTAAGT ATTTGAGTTC CCTCAACTGG GAGAAGGAAA GCAAATGCCC
3851 CTCACCCACT TCCAGTCTATC AATCCACCGG CTGTCACCTT TGAGTTTGTA
3901 AGCCCTTGTT CCTACCGCTC CTGAGTTTCT ATGAAAGGAC CTTGAGGTGT
3951 TCAACAAACA GGAAGGGAT CAACTCTCCC CACCCTGCGT TGACCAATGA
4001 ATTTCTCCCT CCTCTGCTGC CCAGTGAATT AACAGGAGAA AGAACTCCGG
4051 TATTGGAGTT ACCACACATA AAGGATAGTG AGTCAGCAGA GTGCACCCTG
4101 CAGGAACAAT AGAGCCTTCC TTTTCAAGGA AGTTCCTAAGA AAAATGGCAG
4151 CAGGCAGGCC CCACTCGGGT GTATTCATCT ATTCAATTTAT TCAACAAATA
4201 TTTACTAAGT GCCCCTGTGC AAGGCTCGAG GTGTACAAAG ATGAACAGGA
4251 GAGCTAGACT TCTTGCCATG CGTGGTGGGG TTTGCTGCCT AGTGGGAGAG
4301 ACAGACAAAA AGCAAGGAAT GCACACACAG GATGCACACA CAGCGGCAGG
4351 AACCAGGTG CAGTTACCCA GGCCTGGGAT CAGACAGACA GGACTCAGAG
4401 GAGACTTTCC CAGAGAAAAG CCATCTGAGC CAAGGGATGG ATCTGATACC
4451 TCCGAAGGCT GAGCCACCAT AACACTCATA CCTTTAAGCC AAGTCTTATA
4501 AACTCCCAG GTAAGCAGCT GGCAATCAGA AGACCTCCAG CTAATGCCCA
4551 GGACAAGTTG ATGAGCTCTC AAGAAAAAGT TCCTGCCTTT TCTTCTCAAT
4601 ATCCCTGGCA CACAGTTCAG TGAATTTTGA ATGAACCAAT GAATGAAATG
4651 AGCAGGATAT GATAATCCCT CTCCAACACG GAATGTCCAA GCCATGCAGA
4701 GCCGACTGGA AATTTTCCCC GTTCCCTTCC AGATGTTTCT CAGCCAAGAT
4751 CAAAGCTGTG GTGTTTGGGC TGCTGCCTGT GCTCTCCTGG CTCCCCAAGT
4801 ACAAGATTAA AGACTACATC ATTCTTGACC TGCTCGGTGG ACTCAGCGGG
4851 GGATCCATCC AGGTCCCACA AGGTGAAGGG GCTCCTTCAG CCAGGCCTGG
4901 ATTGCCACTC CCCTCACCAT TCCTCTCCTC ATCCCCACTC CATCCCTCTG
4951 TGATCCCATC AAGCTAGTCA TGCTGCTGAG CTTCACTCTC GTTGTCTCTT
5001 GCAGGCATGG CATTGCTCTT GCTGGCCAAC CTTCTTGCAG TCAATGGCCT
5051 CTACTCCTCC TTCTTCCCCC TCCTGACCTA CTTCTTCTCT GGGGGTGTTC
5101 ACCAGATGGT GCCAGGTAAG GCCTCTCCCC TCTGGGCAGG CAGGATGACC
5151 CAGACCACAA GGATGGGAGG TGTGGCAAAG GGGCCTCGGG AGATTTTCCA
5201 TCTGCATTCT CCTGGAGTTG TTCTTGGTCA GTCCTAGGGG AATGGTCACT
5251 GTGAATGTCA TTTCCAGGTC CTCGGTGACC TTGGAGAAAC CACTGAGCCT
5301 CTTTGAGTTC AGTTAGCATT ACCTGTTCCA TCTTCTCTCT AGGAATGAGA
5351 GGAAGACTTA GCAGAACAAG ATATAACATA TGCTATAACA TGCTTAAACA
5401 GATGTGAGAA ATCACCATCT AACTCCCTGG TTGGTCCCAG CCGGCCACTA
5451 CAGGGACATT TGGACTTCTC TGGTGCTAAG TGAGATGGAG GAAAGCCTGG
5501 TCACAAGGGC TGGTTTCTGG TTCAGGCTCT GCTTATATTT CTTATTTCTG
5551 AGTTCAATTT CTCACGTGTC CTGTATGACA ATATTGACCA TTGGGGTAAA
5601 AGCACCTTGA AAAGCATAGA TCATGGTTAG AGTGAGTGGT TGTTATTATT
5651 GTGTTGGAGA AGAGCCTTGG AGGTGCAGGG ATCCATCCCC CTGGGGTCGG
5701 GAAGCATTCC TGGGCCCTT TCTGGTTTCC ATCGGTGTGG TTCAAACCTC
5751 TGATTTTTGC TGGCTGGGTG GGGCACCACA GGTACCTTTG CCGTATCAG
5801 CATCTGGTG GGTAACATCT GTCTGCAGCT GGCCCCAGAG TCGAAATTCC
5851 AGGTCTTCAA CAATGCCACC AATGAGAGCT ATGTGGACAC AGCAGCCATG
5901 GAGGCTGAGA GGCTGCACGT GTCAGCTACG CTAGCCTGCC TCACTGCCAT
5951 CATCCAGGTG AGGGGGCAGC CCCCACCCCT GCTAGAAGGG CATCAGACCA
6001 CCCTGCCCTT CCCTCAAAGC CTTAGCTTTG ATGCTAAATC TGATTTAGGG
6051 GGTGGGTGT GGAGGCTCAT GCCTGTAATC CCAGCACTTT GGGAGGCTGA
6101 GGAGGGTGGA TCACTTGAGG TCAGGAGTTT GAGACCACCT TGACCAACGT
6151 GATGAAACCC CATCTCTACC AAAAATACAA AAATAATCCA GGCTTGGTAG
6201 TATGCGCCTG TAGTCCCACC TACTCAGGAG GCTGAGGCAG GAGAATCACT
6251 TGAATCCGGG AGGCAGAGGT TGCAGTGAGC TGAGATCGCG CCACTGCACT

```

FIGURE 3, page 2 of 20


```

6301 CCAGCCTGGG TGACAGAGCG AGACTCCGTC TCAAAAAAAA AAAAAAAA
6351 AAAAAAAAAA CCCAAGTTAG GGCTCACCTC CTCCCTCCTC CCCATCCCAG
6401 GGCTAAAGTG AACCTTGAAA ATTAACAGTA TCTCCTCATC TGCATGTAGC
6451 AGGACCATAC AAAAAACAA CAGCTGTACC TGGTTAAACT GTCCTGAGCT
6501 TTAAACCTGT AAAAGACTCA CAGCCTCTCT CCATTATCCC GTGGAGAAAC
6551 CCAACTCTCT GCCAGCATAG TCTTGACAGC TGCTAATTTT CTCTAACATC
6601 CCTCACTCCG CTCCAGCCTC CTCTGCTCCA AGCCACAGCA GCAGTTGCAC
6651 AACATAAATT GAGCTTCTGC AAATGGTTGC AAAGGATTCT GCTAGGTTTT
6701 ATGAAGGGAA GCACAACATG ACAGAATGCA AGAGCAAAAC ACAGTCCCAG
6751 AGAGCGCCTT TTCACTCACT CATTCACTCG GTTTTGTGCC AAGAACTAGG
6801 CTAAACCCTG GGATACAAAG ATAAGTAAGA AAGAGGTCCA ATTCACAAGT
6851 TGCTCACAGC CCAGCAGAGG AAGGAGCCAT GTCAACAGAT AAATTGTAT
6901 GCAGTGAGAT AAGCAGCAAA GTAGAGCCAT GTACAAAGAC TGTAGGGACA
6951 CAGAGCAGAG TCACGGAGGA CCTCAAAGAG GAGGTGACAC TCCACCTCTC
7001 TTAAAGGATG AGAACTTAAC CAGGAACAAG GTATACAGAG GATGGTCCAG
7051 GCAGAAGGGA ACAGTGCCTA AAAACACTGA GGCCTGAGAG AGTGTGATCT
7101 CGCGCAGCAA AGTAAGGGGC TTGGTGTGGC TGGAGGGTAG AGGGCCAGA
7151 AGAGGATGGA AAAGTAGGCA GGAGCCAGAC AATGAGATCT GGGGTCTGTT
7201 CTCTGACAGC GACTTTGGGT CTGATTGGCA GTTTATAAGG ATCGTTTGGG
7251 CTACACAATG ATGAGTGGGA GGTGGATTAG AATCAAGGCA GGGGACCTGT
7301 TGGGAGACTC TGCAGAGGCC CAGGCAGGAA TAATGCAGGC GAAGACCAGG
7351 TAGAGAAAGA GATGGGGCTG GACTTGAAAA GAATGTTTTA CCAGGAGCTT
7401 GGTGATAGAC TGGATGTGGG AGGTAAGGGA GGATGACTCT CAAGTTTTTG
7451 GTTGGGCAAC CAGTTAATG ATGGTGTCTT TACTGAGAG AGAAAACACT
7501 GGGGGAGGAC TAGACTTATT TTACAGATAA GCCAAAGCCA GAGAGGTGAT
7551 GTGACAGAAA GGCCATGCT CTAAAGGAGC TGAAGGTCTG ATGGCAGCCA
7601 TGTAGAGCAC AGTGAAGGGC AGGTGAAGGT CACAGATGGT CCAATCCCT
7651 CAAGCTACTG CTACGCTAGG ACTGCACGGA GCTCCAGACC TGCCTGTGTG
7701 TGGGGCGGGT CGTTGGAAT GCTGAACCAC ATTGGTCTTC CGCCACCAAC
7751 CACCCTTTTC CTCTCTCAG ATGGGTCTGG GCTTCATGCA GTTTGGCTTT
7801 GTGGCCATCT ACCTCTCCGA GTCCTTCATC CGGGGCTTCA TGACGGCCGC
7851 CGGCCTGCAG ATCCTGATTT CGGTGCTCAA GTACATCTTC GGAAGTACCA
7901 TCCCTCTCTA CACAGGCCCA GGTCCATCG TCTTTGTGAG TCTGGGGATG
7951 CACCCTGCC ATTGGAGCAA GGCTCCAGCA GACACATGAG GAGGATGTAC
8001 TGTTTTAAGA TGTCGTGAGC TCCTCATTGC AAGGGCTGGC TTAGCTGTG
8051 TTCAGAGAGG ATTCTGAGGG GGTTCGTGTC TTGGGAGGGT CAAAGTCATG
8101 ACTCACAGAG GTTCTTGTA GTTAATACCT GCAGAAAAGA GCTGTACATT
8151 CTCGCCCAGT TCCCCATTCT AGTGCCTCAA CCCCTCCCTG CCTGGAAAGT
8201 CCTGCCTTAT GTCTAATCTC CATCCCTCCT CCTTCAGCCC AAACCTCTCT
8251 AAAGAAAAAG AAAGCATTCC TTTTCTAGCA CAAGTTCCCC ATGTGCCTTT
8301 TGGGAAAGGG CGGTGGGCGA CGGGACAGGG TTCCTGATCA GGGTTTTAAT
8351 TCTGTCTTGG TGTGCCTCCA TTAGCTTTGA TGGCATCCCT TCCCTGGGTC
8401 AGACACCCAA AGGTGGGTA TTATGGGAAG AAGGGGTGGG AGCCTGTGAG
8451 CATGATGCTC TTTCCCCAG ACCTTCATTG ACATTTGCAA AAACCTCCCC
8501 CACACCAACA TCGCCTCGCT CATCTTCGCT CTCATCAGCG GTGCCTTCCT
8551 GGTGCTGGTG AAGGAGCTCA ATGCTCGCTA CATGCACAAG ATTCGCTTCC
8601 CCATCCCTAC AGAGATGATT GTGGTAAGGA CCTTGTTTCA AGCTGGGATG
8651 TTGGGGGGCC AGGCTGTGAG ACGAGGAAGC CCCTACCTTT CCTCACCCCA
8701 TCCCTCAAC TGGCAGCCAG TGGGACAGGA AGTCAGTTGT GAATCCATCC
8751 CATCCCCCGT ATGTGGCGTT TCCTCTCTTT CTAAGTCTCT AATAATTCCC
8801 CCTAAGGAGG CAGGGGAGTG GGATTCAGGG TCCCAGAGA AAAGGGAGAC
8851 TTGAGAGAGA CGCCTGCCCT GGCCCCACCT TAGGGCCAAT CCCATTCTC
8901 CACTCTGGGG TTTGACAGGT GTGGTGGCAA CAGCTATCTC CGGGGGCTGT
8951 AAGATGCCCA AAAAGTATCA CATGCAGATC GTGGGAGAAA TCCAACGCGG
9001 GTGAGTCCAG GTGGCCCAAG AGCCTGGCCC ACCCGCACCT CATGCCCCAC
9051 TAAGGCCTGA GCTCGGAGAG GGAGACAAGA TGAAGTCTAT GAAAGTGCAG
9101 TCGAAACTGT ATGACACTGA CCATGTATGA ATTATTACTA TTACCGTTTC
9151 CTGAGAAGGG CCGCACAACC AGCCAATGTA GGCTATTTTA TGAGAAATGA
9201 GTCTTAACTG CCACACTCCC CTTATAAATC TCATTCAACT GATGCTGTTA
9251 AACAAAGCCT CTCTGAACAG CCGCTTGCTG GCTCTTTGCC TTGCTCTAAT
9301 GCATTGGTTC TTTGTCCATG TAGAAAGGGA ACTATTAGGT TCAACCAGAT
9351 TCATGAAGCA TCCACTCTGT GCCAGGCACC ATGCTGGGCC CTGGGAGGAG
9401 AGGGGTGACG CTTGTCTCTG AGGGTTGGAA CAGGCAAGGG AGGGAAGACC

```

FIGURE 3, page 3 of 20

9451 ACATAGCACC AAAGGTCTAG GGGTCTGTGG ACTCGTGAGC ATACAGGGTT
9501 CAGAATCTGG GAGTTAACAA ACGAGGCCCT ACCACATACT GGCCCGGGGA
9551 CCTTGGGCAA GTTAGGTTCT CTCAGCCTCA GTTTCCTCCT TTGTAAAACA
9601 GGAGTGATGG TCCCTACCCT ATGGGGTGGT GCTGAGGATT CAGACTGGAT
9651 GGGATAACTT AGGCAAAGAT CCCGGCACAC CATGGGGGCC TGGCTGGTCC
9701 CTGTGGGCTG GTGAAGGACT TGGCTGCCCT CCCCCTCAC ACCCTTGGGT
9751 TCTGCCCTCT TCTGGGCTCC TCGGCAGGTT CCCCACCCCG GTGTGCGCTG
9801 TGGTCTCACA GTGAAGGAC ATGATAGGCA CAGCCTTCTC CCTAGCCATC
9851 GTGAGCTACG TCATCAACCT GGCTATGGGC CGGACCCTGG CCAACAAGCA
9901 CGGCTACGAC GTGGATTCTGA ACCAGGTAGC TCTGGCCACC CCCGGCAGGA
9951 CTGGGCAGGA CAGGTCAACT CAGGCCTGGC ATGACATATC TTGGGTGGGG
10001 AGATCATTGG GCTGAGGTGA GGCAGGCTGC CTCGAGTGTG GGGGATAGGG
10051 GGTCCTCTGA CCCTAAGAGG CTGACCTCCT CTTGACTGGG AATGTGTGAC
10101 TTTATAGCCA CTGGGTCACT CTCAGGTCTT AGGCCACAG TCCAGCTTGC
10151 ATGCCCTGACT GCACTTGGTC CCCGTGCCCC CCAGCCCCAC ACTGGCTTCT
10201 AATCCTGTCT CCTCCCTGCA GGAGATGATC GCTCTCGGCT GCAGCAACTT
10251 CTTTGGCTCC TTCTTTAAAA TTCATGTCAT TTGCTGTGCG CTTTCTGTCA
10301 CTCTGGCTGT GGATGGAGCT GGAGGAAAAT CCCAGGTGAG CTTTGTCTTA
10351 GGGGAGTTGG GGGGAGGTGG TAAGAGAACA GTTGCCCCAA AAAAGCCTGG
10401 GCACTGCAAG CCAGGCCAGC TCTTCTCCGA CCCCTTCTTC CCGTACTTAG
10451 TCTCCACTCC ACCAAGCCA TGATTGGAA ATAAATCAAG AGCAAAAATT
10501 TCACACCTTC CCTCTATCCC CAACTCTTTC TCGGAATAGG TGGCCAGCCT
10551 GTGTGTGTCT CTGGTGGTGA TGATCACCAT GCTGGTCCTG GGGATCTATC
10601 TGTATCCTCT CCCTAAGGTA AGAGCCCAGC CATCGAGCAG AAGTCAACGA
10651 AAGACTCCAA TAAGAACAAT CCCTGAGAGT TGTGTGGCAC TTTACGGACC
10701 ACAAAGTGCC ACTGTTGTCA TACTTAGTCT CAACCACAAA CTGTGAGGTA
10751 GACAATGCAG GTTTTATCCT CCCATTTTTC CAGGTGAAGG AAAGTGAAGT
10801 TGAGAGTCTA AGTAACCTTG TCCATAGTGA GGCAGCTTAC AGCGCAGGGC
10851 TGGTCCCAAA CTCCAGCCTT CTGGCCTCAG AGTCTAATCC CTAGGCAACA
10901 TTTGCACCTA CCCACGAGTA CCAGGCTCTT ATATAGCCCA GCTAGGAGGG
10951 CTCTAGGCAT CCGTCAATTA GAGATGAGGG AAGAGAGATA GGGAAAGGAT
11001 GGGGCCAGGA AGGACCCCAT GGCTCTAACG CCAGCACTTT CCAAACCTAA
11051 GGTCGAATGC AGAGATTTGG GGGATCAGCC AGGGGAGGTG TTCCAGAAGT
11101 CCGTCTCTGT CCTGCCAGGC CTTGGGGTCG GGTATGCGCA GGAGGGCAAA
11151 AAGAAGGGGA GACCCTGGGG TCCTGGAGCA ATGTTCTGCT TCTCTAGTCT
11201 GTGCTAGGAG CCCTGATCGC TGTCAATCTC AAGAACTCCC TCAAGCAACT
11251 CACCGACCCC TACTACCTGT GGAGGAAGAG CAAGCTGGAC TGTGTAAGTA
11301 TCGGGCAGCC TCTGGGTACT GGCCATGCCC CTGCCCTCTC CTCCAACCCC
11351 ACAGCCCTGT CAGCCCTGTC CTAACAATGA ACCCTCTAGT CTGCTGCTTC
11401 CTAATTAGCA TGAGATGAGT GGTAAAGT CCGAGTTTCG AAGTGAAACA
11451 TCCTATGTTT AAACCCCTAAC TCAGCCATCT GCTGGCTCCA TGGCCAATAG
11501 CAAGCCCCTT AACCTTTCCC AGTCTTGGTG TCTTAACTGG GCAAATGGTT
11551 ATTTTATGCT CTCTGCCCTC CAGGGTTTTC TATGAAGAAG AAGCAAGGTA
11601 ATACAAGTAA ACATGTTGTC TACATCGTAT TTTATACTCA ATAAAGCTTA
11651 GCTATGACTA CTTTATGACA TACAGCTTTA AAAACAAAAA GGAAATAGTT
11701 TGTATTTTAA AAAAAACCT AGAACATAAA GCCAGAGGAC CAAAATCTTG
11751 AGCAAGTTAC TAGACTTCCC TGGGGTCTA TTTCTCATC TGTAATGGG
11801 GGTGAGACTC ATGCAGTCAT GGTGCGTCA AACGCTGGT CCGAGGATTA
11851 AATGAGATCC CAGTGGGAAA ACACCGCATG AGCGCAAACA TTCTGCAAAAC
11901 ATGACTTATT GTCCTGATTA GTCACACACT CCACCGCATC ATCCGCTGGG
11951 CATAGTAATG AAGGCCAGTG TGTTTTGACG AACTGCCTT CTCTCCATTT
12001 AAGCCCCACC ATAACCTATG GGAGAGGATT TACTAACTT TCTTAACGGT
12051 GATGAAACCA AGGCTCAGAA TGGTTAAGTA AATTGTCAAA GGCCACAGAG
12101 GTAGGGAGTG GTAGAGTCTG GATTAAAACT CCAAGTCTTG GACTCCAGAC
12151 CTCTAGGCTG TACTGTCTCA TAGGGAAGGC AGTCTCACC ACCTAGGGCA
12201 GAGAAGAAAA TCCTTAAAGC CAGAGAAGTG AGTGGTCTAT CTGTGGTCAC
12251 CCAGAGAGAC AGTGATGAGG ACAGGGAGAA AAATTATACC TCAGTTCCCA
12301 GCCCAAGGAT CTGCTTTGAC CATAACCCAA CAAGCCCCCG CTATGGTGGT
12351 ATTTCTTTAG GTTCATATGG CGGCTTTTGT TTCCATTGTA TCTTCACAGC
12401 AATTCTCTAC AGGAATCTGG GCAGATTTAT TTCCTTTAGA GGAATTTCCA
12451 GGTCTTAAAA TCTATAGGGG GCAACTATCA AAATTTACC CAATGTTGCC
12501 CCCTACCCAC ACACAAAACC AGGCCCCCAG CCGATCAGAA AGCACTGCTG
12551 AGCTCCTGTC AGGGCCCCAG CAGCTCGCTG TGAGACAGAG AGAGGGAAC

FIGURE 3, page 4 of 20

12601 CACATTTATT GATCACCTAC TGAGCATCCA TCACTAGGCT AGGACCGTCA
12651 CATTCCTTAA CTTTTGAATC CTTTCATGAG GTAGGCATTA TTATTCTCCT
12701 TTTGTTTCAC ATAGCCATTA AAGAACAAAA TTTGGGGCTG GGTGTGCTGA
12751 CTCACACCTG TGATCTAGCA CTTTAGGGGG CTGAGGCAGG AGGATCGCTT
12801 GAAGTCAGGA TTTCAAGGTC AGCTTGGGCA GCTTAGCGAG AGCCGTCTCT
12851 AGAAAAATAT AAAAGTTAGC TGGGTGTGGT GGCACGTGCC TATAGTCCTA
12901 ACTATTCAGG AAGGTTAGGC GGGAGCACAA CTTGGGTTC AGGGTTTGAG
12951 GCTCCAGTGA GCTGATCTTG CCACTGCACT ACAGCCTGAG CAACAGAGCA
13001 AGACCCTGTG ACTCCAAAAA CAAACAAACA AACACATTTT GAACCCAAAC
13051 AGATCTGACC CAAGATGCAT GCTCTTATAG ATGCCACCTC CCTGTGTGCT
13101 GGGGCTTCTA CTA AAAACAC AGACAAGATC AGGCAACCAC AGTCAATCTA
13151 AGGGAAAGAG GAAAGTGTAA CCAAAGCACA AATACATAAA ATATTGCAAA
13201 AATGCTATTT AAAGAAAAAA AAGAGAAGAG AGGCTCTGAG GTTGTACTAA
13251 CAGAGAATGG CTTTGGCTAA TCCAGGAAGA CTTCCTGAAA GAGGTTGTTT
13301 TTTCCCCAGG TCTGCTTTTG ACATCTCTCT TTTACAGTG CATCTGGGTA
13351 GTGAGCTTCC TCTCCTCCTT CTTCTCAGC CTGCCCTATG GTGTGGCAGT
13401 GGGTGTGCGC TCTCCGTCC TGGTCGTGGT CTTCCAGACT CAGTTGTAAG
13451 TGATAGCTTC CGCCCTCCTA GGCCACAGT CGGTTCCCTG GGCCAGCCCCG
13501 CAAAGGGCTT CCATGCCACG GCCTGGCTTA GTCCACTGTA CCTTCCACCT
13551 CTGGGCCTGG CACTGGAGGT GCTGCCAGGC CCAAAGAGAG CCCAACCCAG
13601 CCAGGACTGT GGGCACAGTC TGGGCTGTTT GACTTCCCAT ATCTTGAAAA
13651 CCCAGAGAA AGCCAGCATA CTCTTGCTGG GGATGGCTGG GGAGAGGGCA
13701 GTGGCAGAGA AAGGAGGGCA AGGGCAGGTG GTGAGATTCA ACATCCTTCC
13751 AAAGACATTG CCAGAACCCC AAACCAATG GGACCCACC CCAGGAGAGC
13801 GCCAGGGTGG AAGACAGAAG CTGTGTTCTA CACACTGGGA GTATTACAGA
13851 GAAGGGGTCT TGGCCAAGGC AGGGAGTACG CTGAATGTTG GGGGAATCCT
13901 ATCTTCTCTT CTTGAGAACT CAGAACAAGG AAATGATGAC TTCAGGGCGA
13951 CTCCCACCAC TTCTCCCACC ACTTCTCTCC CCTGCCCTGT GGTCTGGGAG
14001 CTATGTCAAG GACCTGCCTG TCATCCTCAT AGTTATAGGA GGCCACAGGC
14051 CACCAGACAT GTGTCTCCAG TGCAAAAAGA CAGACACAGC AAGTCTGGGG
14101 GTGAGGACAG GACCCCATCC TACCTTGGCT CTGCCCCCGC CCCAGCAGGG
14151 GCACCCCTCC AGGCCCATGT GCCATTAGCA TTCTCTTATG TTTTCTCTT
14201 CCTGCTTCAT CCAGTCGAAA TGGCTATGCA CTGGCCAGG TCATGGACAC
14251 TGACATTAT GTGAATCCCA AGACCTATAA TAGGGTAGGT AATTCAAGCT
14301 TATGACCTCC TTCTTTTGCT CTGCACCACC CCAAGAAGAG GTTGCTTTTT
14351 AAAGCCAATA AAGACATTTT TGCAACTTGA GCTCAGTCTC CCTGTACAG
14401 GCCCAGGATA TCCAGGGGAT TAAAATCATC ACGTACTGCT CCCCTCTCTA
14451 CTTTGCCAAAC TCAGAGATCT TCAGGCAAAA GGTATCGCC AAGGTAAGGC
14501 TCAGTCCCTG GCGACCAGAG GCTCTGGACA GAGAGTGGCC GGAAAATGGA
14551 AGCAGAAGGG CGGTGGGAGC TGAGAATAGG CCACTCCCAT AGAGGGTGGA
14601 GGTCAAGATT GCTGTTGGCT CTCTCCCTGC AGACAGGCAT GGACCCCCAG
14651 AAAGTATTAC TAGCCAAGCA AAAATACCTC AAGAAGCAGG AGAAGCGGAG
14701 AATGAGGCCC ACACAACAGA GGAGGTCTCT ATTATGAAA ACCAAGGTGA
14751 ATGAAGGCCA GAAGCAGCCC CGTGCCTGC TCTCCTGCCC ATTCTGATAC
14801 TGCCCCCTGT TACTCATGGT ACCCTGGGGG CCCCCTTCC CACCCTGACA
14851 GGCAAAGACA GAAAGTCTCT GGGAACTCTG CCTGGTGGCC GCTGGGCATT
14901 TTTCTTCTTT TTTTCTTTT TCTTTTAGA GATGGAATTT TGCTCTTGTC
14951 ACCCAGGCTT GAGTGCAATG GCGTTATCTT GGCTCACTGC AACCTCCACC
15001 TCTGGGGTTC AAGCGATTCT CCTGCCTTAG CCTCCCAAGT CGCTGAGATT
15051 ACAGGTGCCA CCACACCCAG CTAATTTTGT TATTTTGTAGT AGATATTGGG
15101 TTTCACCATG TTGGCCAGGC TGGTGTCAAA CTCCTGACCT CAGGTGATCC
15151 ACCTACCTTA GCCTTCCAAA GTGCTGGGAT TACAAGCCTG AGCCACTGCG
15201 CCCAGCCTGG GCATTTTTCT TCTTGGATGA GGTGCTACCA TCTCCAGGG
15251 AAGCCACTGA ACCCCCAAGG CCCTTCTCCA TTTTCTGGCT AAGATAGGAC
15301 ATGGCCCATG GACTTTTGAA CAACCCAGAG GGGGAACAGC AGTGAATTTT
15351 CTGGGGAACC CAGGCAGCCC AGGGCTAGCA AGGCTGGGGT GGCCATGGCA
15401 GTAATCCTTG TAATCCCAGC ACTTTAGGAG GCCGAGATGG GAGAATCACT
15451 CTCATGAGTT CAGGAGTTCC AGACCAGCCT GCCCAACGTG GCGAAACGCT
15501 GTCTCTACTA AAAATACACA AAAATTAGCC AGGCGTGGTG GTGGGCACCT
15551 GTAATCCAG CTAATCAGGA GGCTGAGGCA CGAGAATCAC TTGAACCCGG
15601 GAGGCAGAGG TTGCAAGTGA CCGAGATAGT GCCACTGCAC TCCAGCCTAG
15651 GCAACAGAGG GAGACTCTGT CTCAAGAAAT AAAGGAGCTC AGTGTCCCCG
15701 GAGGGGCTTT CTCCCAGAGA GAGTGGGCTT GAGGCTTCAG TGCCTCTCTT

FIGURE 3, page 5 of 20

15751 GGCTGGGTCC TCTGACTTTG TCTGGGTTGT AGGAGACCAA GTTTCAGGC
15801 CCTGCCTAAG AAAGGGCTTT GGGAGAGGCC TCTCTGGTGG AGCTTTCAGG
15851 GTCTGTGTTT ACCATCACCG AGGCGAGTTA TTCCCTTACA CCTACACCCT
15901 CCATGCCCCCT GCTTCAGTCA CAGCAAGGTC TGGCTCAGTC TGGTGGTCCC
15951 TGA CTCTGCC CACTGTCCCC ACCCTTCCAG ACTGTCTCCC TGCAGGAGCT
16001 GCAGCAGGAC TTTGAGAATG CGCCCCCAC CGACCCCAAC AACACCAGA
16051 CCCC GGCTAA CGGCACCAGC GTGTCTTATA TCACCTTCAG CCTGACAGC
16101 TCCTCACCTG CCCAGAGTGA GCCACCAGCC TCCGCTGAGG CCCCCGGCGA
16151 GCCCAGTGAC ATGCTGGCCA GCGTCCCACC CTTTCGTACC TTCCACACCC
16201 TCATCCTGGA CATGAGTGGA GTCAGCTTCG TGGACTTGAT GGGCATCAAG
16251 GCCCTGGCCA AGGTGAGGCC CTCGGGGACA GCAAGCACCA CCCACTCCAC
16301 CCCCTCCGCT CTGCTCTCCA CATTCCCTTT CCTGGGAGCC CTCATTTCAG
16351 GAAGCTGAGG GAGGAAGCTC ACTGGGGAGA CTAACAGCTC CTAGGAATCC
16401 CTCCTTTCCC CAGACGCCAC CAGGTTGAGA CATTCTCCAC AGAGCAGGCC
16451 CAGACGGCCC ATGACAATGA GTGGCGGGAC AAGTCTACCA GAGTTTCAGG
16501 CCCCTGTGCT CCCAACACCC CCAGCAGTGG CCATCCCAAG TCCCTCTCAG
16551 CCATCAGGAA CCCACCCAGG TTCTCTGAGG AGGGTCCAGT TTGGCTCCTG
16601 GTTCATGATC TGGTGCCCTT GTCCCTCATT CACCAGCCAC CCTAGGACAG
16651 GAGAAGAAAT AATACCAGTG CCCACACCA TCAGGCCAAA CAGAGAGCCC
16701 ACGGGACACC TTGAATGAAT GTATCCATCT GATAACTTTC CAGCAGCCAC
16751 CGCCAATGGC GGGAGTCAGC AAACCTCAGA GCTGGCTCAG ATAGAGGCAA
16801 GCCAGGGGAA CAATGGGCAC AGAGAGTGTG CGGACTGCCT TCACCATCAA
16851 CCAGGCGCAG GGCAGGCCCC ATACCCAGCC TTGGGCTCA GCCGGCTTCC
16901 TTAGCCAGGA TCTGGAGTCC AGGCCAGCCT TGGCTGAAGC TCTAGACTCC
16951 CTGAGCCTCC ATCCTCCCCT GCAGCTTCTG TCTGAAGCCA CAAAGAAGTC
17001 TGAGAATCTA AGCTACTGAA AGAAAAGATC AGCCGGGCGT GGTGGCTCAC
17051 TCCTGTAATC CCAGCACTTT GGGAGGCCAA GGCAGGTGGA TCACAAGGTC
17101 AGGAGTTCAA GACCAGCCTG GCCAACATGG TGAAACCCCG CCTCTACTAA
17151 AAATACAAAA ATTAGCCAGG TGTGGTGACG GGCCCTGTG GTCCCAGCTA
17201 CTCGGTAGGC TGAGGCAGAG AATTGCTTGA ACCCAGGAGG CGGAGGTTGC
17251 AGTGAGCCAA GATCGGCCA CTGCACTCCA GCCTGGGCAA CAGAGTGAAA
17301 CTCCATCTCA AAAGAAAAAA AAAGAAAATA TCTAGCCCCA CAAGAAGGGG
17351 CCATGGTGAC TTTAAGTGCC CGCCACGTTG GCAAAAGTCC ATTTCCGCTC
17401 CACTTCCAG AGAAACCGTC AGCCAACACT CCAGGGAGAA GTGGTGTGCT
17451 TTGCTGCTAT TTTTGTCTTT GGCTGCTGGG CTCTCAGGGT TGCTTATTTG
17501 TTTGGCTTCC CCTCTGAAGT ACGTTTTGTG AATCACTTTT GAGACCCACT
17551 CAGAACATTC CTTTCTTTT GCCTCCCTAC CCCAACAACA CTTCTAGCTG
17601 AGCTCCACCT ATGGGAAGAT CGGCGTGAAG GTCTTCTTGG TGAACATCCA
17651 TGGTAAGAGA AAGAGGACAT TTAGGGACTG AAAGACTGGC AAGGAGTGTG
17701 GGGTAGGAAC AGGTTGGTGG GGTCTGAATA GTGAGGAGGT TGGAAACGAG
17751 AGCACCCAGC TATCCCCCAC AAGCTGCTGC CTGCTCATAA AAGCTTCAGG
17801 TACAAGTCCA AAGAGACTGG TCAGATTGCA TAAACATCCT AGGGGCCTTA
17851 GTGACAGAGT GGGGGTGAGG AGGTGATGGA GTTACAGAAG GACAGCTAGG
17901 ATTCTAATCT ACCCCATAAC TAATTGCGA CGTATCCTTG GCCGAGTCAC
17951 TTTATCTCTC AAGGGATCTA TTTCTACCTA TGTAAAACGA GAGGGTTGAC
18001 TAGATGGATT TGGGGATCCT CTCCAATCA GAAACTCTGT GAATCGATAT
18051 AGGCATAGAG CACACGGTAC CCTAATTTCC CAGGGAACAT ATAAATATGC
18101 AGTTTTGTAG GCATACAGCC TCCAAAGGGT GCATATACAC AGCCTCAAGG
18151 ACGTGGCCAC AGGGCAGCAG ACATTTACAT GACTAGCATG TACGCAAAGT
18201 GCAGAGATGT GGGAGCAAAGT GCACACAGAC ACACAGGAGA ATGTGAAGGG
18251 GCACATACAC ACACACCCAG CTCCCTGCAC TGGGTGAGC CCCCTCCAGC
18301 AGGGCTGCAG TTCCCAAGCT CCGCATGGCC ACGTTCGGGG AGAGAATCTG
18351 CAGTGGCAAT GACCTGCTAT GATATGTTCT GGAGTTAGAA GCAGTGGATT
18401 CTCCTCAACC TCACTGGACA CCCCCTTAGG AAACCATCTC TAGGATTAAG
18451 AGTAATCCAC ACAAATCTCC AATGCCACAC ATTGGAAGTT GCTGGAAGG
18501 TCTGGGAAAA CAAGAGGAAG GATGGGTCTT TGGGGGATAG AACTGGCAGC
18551 GGCCTCTTCA AGGATGGCTT AGGCTTTTCC ACTCGAATCA CCACAAAGTA
18601 CTGACTCCCT AAATCAAAC TCTTCTTCT GCTCTGGGT GAAACTTCAG
18651 CATCCTCAAG TTCATGTTGC CCTCTGCCGT CCAGAACTGA TATTGCACTG
18701 CCAATGCCAT GCCCTCAGA TACAGCAAGA GCTGGGACCT CAGGCTCTC
18751 CCAATCCCTGC TCTGGTCTCA CTATCTTCCC CACCCCGAGC TCCAATCCAC
18801 AATGGCTGTT ATCTTTCTGA AGGTGATCTT TTCTCCTTCT AGCCAGGTG
18851 TACAATGACA TTAGCCATGG AGGCGTCTT GAGGATGGGA GTCTAGAATG

FIGURE 3, page 6 of 20

18901 CAAGCACGTC TTTCCCAGCA TACATGACGC AGTCCTCTTT GCCCAGGCAA
18951 ATGCTAGAGA CGTGACCCCA GGACACAAC TCCAAGGGGT AAGGTTCTTG
19001 CACCTGGGGA ATCCTAGGCT CCAAGGCACT GAAATAGCAG GACCAAGAGG
19051 CATTATTAGA AAGAACACAG GAGAAGGTTT AAGTTCCAAT ATCAAGTCTG
19101 CCATTTTCAGT TTTCTGAATC TGTTTCCTTA TCTATAGAAT GAGCACCATC
19151 AACTAACATT ACCTACCTCT CTGCATTTTT CTTTTATTTT GTTTTAGGGT
19201 TAAATGATAA TTACATCTTT TGTGTCACTT GAAAGCACTT TGTGTATTGT
19251 AAAAATTCTT TATCAATATA AGTTTTCTGG TTGCACAAAC ACCCAAAGCA
19301 TAGTAGAGCA GGCCCACTCT GCTGGCATCG TTCCCTGCCT CCTCCTCATC
19351 TCTTTCTAAA GGGGCTTTT GGAAGGGAG GGGAGGGAG TAAGCCTACC
19401 CATTTTAAC TACCGGAGCT TAGAGATTTT AGGCTGGTGA GGGATAAAGA
19451 GATTGGGTCT GAGTTTTGTC TCAGCTTTTT GACATTTAAT TTAGTAGCTC
19501 AGTAAGTCAT ACAAATGGGA TACAAATAAC ACCATCTAAA ACTCCAGAAG
19551 ACTGGGGAGT CAGAAAAATC CTACCTCCTT GGGGTCCCTG CCCAGATCCC
19601 CAGTCATCTC TAGCCCTCAG GGTCCCCTCC CAGCTCAGCT CCTGCCCTTG
19651 GCCTCCCAAG ACTCTTGTG TGCCCCAGCC CTGGGTAAAA ACCTCCCCTG
19701 CCCTCTGTGG GTCATAAGAA AGGCTTTTCT GGCCTTAGAG CAATGATTTG
19751 CTCTTTGCCT TAAGAGACTG ATGAAGGTGA AACCATCTGT TCTAAGTGCT
19801 GAAAGACTGC CCAGGAACAC ACAGGGCGCT GGCTCCTGCC CTCCATGCCT
19851 AGAGGGAAAC CCTGGGAAA CAACGGGCTT TCCTGCTTCG TGAAATTGT
19901 CCGCAGAGCA AAGAGGGAGA TTCTGGAGGA AGCTGCATTA GTTGTAGTG
19951 CCCTAATCAT GTTCAGCTAC TCTAGTTGGT ATGTATACTT GATTAGTCAT
20001 AGCACTTATA AATAATTTAT ATTTTATATA ATATATACTT ACATATTATA
20051 GACCATTAC AGATACAAAT CACACACATA AACACACACC TTTTCAACAG
20101 CATTGTGAGG GACAAAGCAG GCAAAGTGAG GCTGGTTATC AGACTTTAAC
20151 AGATTAGAAA ATATATTCCC AGGAGGACAG GAATTCCCCA AGGTCAGGCA
20201 GCTAGCCAAT AGTTTTTCTA AGCTGAGTAA AACCTTCCCT GCCTCTAACG
20251 GCCACAAAG GAGGGAAGAC CGCGATACAC ACCTGTCTGG TATAAGGGGG
20301 AAGACCACAG CCGTGCTGTT TTTGTGAGGC AGGTAAGGGA AGGGGCAAGA
20351 GGATAAGTCA TGTGTCAGGA AGCAGCGTCC AACCAGAGCC GGCCACCTGT
20401 CCCTTTTCTT GCCACCATGC ACCAACTTTG CTGTTCACTC ACTGAAGCTC
20451 ATTCTGCACT GGCTTCTCTC CTTCAGGCT CCAGGGGATG CTGAGCTCTC
20501 CTTGTACGAC TCAGAGGAGG ACATTGCGAG CTACTGGGAC TTAGAGCAGG
20551 TGAGCTGAGG GAAGGGGCTG TGAGGGTGGG AGCAGGGCGA AGAGGGGAAG
20601 GATGGGGTCG CTGTCAAATA CAAGGCGTTC ACTCAGCTGT CTCACCTCCA
20651 GCCCAGAGCA GTCACATTCA AGGCCACAAA GATTGTGGT CATCTTTGTT
20701 TTTTCTCTTT TCCTTTTCTT TTTTCTTTT TTTTAATTTG AGACAAAGTC
20751 TCACTCTATC ACCCAGACTG GAATGCAGTG GCATGATCTC AGCTCACTGC
20801 AACCTCTGCC TCCCGGGTTC CAGAGGTTCT CCTGCCTCAG CCTCCGAGT
20851 AGCTGGGACT TCAGGCCTGC GCCCAGCTAA TTTTGTATT TTTAGTAGAG
20901 ACAGCTTTTC ACCATGTTGG CTGGGCTGGT CTCGAACCTC CGATCTCAAG
20951 CAATCTGCCT GCCTCGGTCT CTAAGTGCC TGGATTACAG GCATAAGCCA
21001 CGATGCCTGG CTTTGTCTT CATTCCTCTC ACTCCCTGAA AGGCATCGTG
21051 GGGAGAGGGT GAGTCACTGG ACCAAGTCTT AGAGAACCAG TATCTATTCT
21101 TATTCTCCAA CACATCACCC ACGTGACCCT GAGCAAGCCA CATAACCCT
21151 GGGCCCTAGT TTTTATCATC TGTGAAATTA GGGGAAACAT AGGTAATACC
21201 TGTCCCATCC ACCACACAAG ATTGGCAGGG CAGTCACTTG TTCTTTTATT
21251 AATTCAGCAG GTATTTATGG CGTACCTACT GTTGCCTGA CACAGTTCAG
21301 GATGGGCACA TAGCAGTGAG CAAAACAAAG GCCTCTGCCT TTTAGAAACT
21351 TACGTTATGG TAGAATAGAT GGATTTNNNN NNNNNNNNNN NNNNNNNNNN
21401 NNNNNNNNNN NNNNNNNNNN NNNNNNGTCT ACAAATGAAT TATTATTGCA
21451 TGTGGACAAG CCTTAAGAAC TAAAAAATAT GTGGCTGGGT GCAATGGTTC
21501 ACACCTGTAA TCCAGCACT TTGGGAGGCT GAGGTGGGCG GACCACCTGA
21551 GGTGAGGAGT TTGAGACCAG CCTGGCCAAC ATGGCGAAAC CCCGTCTCTA
21601 CTAAGACAC AAAAATTAGC CAGGCGTAGT GGTGCATGCC TGTAGTCCCA
21651 GCTACTCGGA AGTCTGAGGC ATGAGAATCA CTTGAACCTG GGAGGCAGAT
21701 GTTGCACTGA GCCGAGATCG TGCCACTGCA CTCCAGCTTG GGTGACAGAG
21751 CTAGACTGTC TCAAAAACAA ACAAACAAAA CAAAACCTAA AAGATATGTG
21801 GATATGAGGG ATCACCATCC CCATAGGGCC CCTGGATTAA CACCACCCCA
21851 CCAATGCCCT GAATTAAGAG AAACCAGATG ACTAGGTTTG GAGAAATCTG
21901 GCTTTGGGTC TATGAGAAGT AGTGTCTCTC TTTGTGCCTC TTCCCATCTT
21951 TTTTGACATT GAGCTCCATG GTGCTCTGAA TCCGTCTCTC ACAGTGCTGA
22001 TGGCAGGTGG GACAGATTAG AAAATAGAGC TGGAGCCACA GAGATTGGC

FIGURE 3, page 7 of 20

22051	AGACTGATTT	CGGTGCCCTC	TTGGAATCTC	CAGCACATTC	CAAAAAGCCT
22101	GGATAGGACC	AAAATAGCTT	ATCAACGTGA	GAAAGGACTT	CAGAGCTTGT
22151	CTACTGCCAA	CCCTCATTTT	ACCCATGAG	GAAAGTGAAG	CTATTAGGGG
22201	GCGAGGGACA	CGTGGAAGGT	CACACAGCAC	ACAGGAGGTG	ATTACATATG
22251	AGATTTTCAG	ACCTGCTCCT	GCCACGCTGG	ACTGGTTTAC	CTCCTAGGCT
22301	GACCCCTGCC	CTCCCCCTCT	CACACACAGT	CTCGGCACACA	CACACACACA
22351	CACACACACA	CACAGGTGCT	TTGTTCTGGC	CAGGGGTTCC	TAGGGTCACC
22401	TCTTGTTTGC	AGCCACTGTG	ACCCCAACTG	GTCTAACCTC	TCTCTTCCCC
22451	TCCCACTTCC	TTCTGTGGT	TCCTGCAGGA	GATGTTCTGG	AGCATGTTTT
22501	ACGCAGAGAC	CCTGACCGCC	CTGTGAGGGC	TCAGCCAGTC	CTCATGCTGC
22551	CTACAGAGTG	CCTGGCACTT	GGGACTTCCA	TAAAGGATGA	GCCTGGGGTG
22601	ACAGGGGGTG	TCGGGCGGAG	GAAAGTGTCAT	CCCCAGAGCA	TTGGGTTCTCT
22651	CTAGCTCTCT	CCCTCTCTCT	CTCCCTTCTCT	TCCCTCCCCG	CATCTCCAGA
22701	GAGCCCTCTCT	CAGCAGCAGG	GGGGTGCTAC	CCTTACAGGA	GTGAGAGTCT
22751	GGTGAGCCCA	CTCTTCACCC	GTGAGGCCCT	GGCCGCAATG	GACAAGCCTC
22801	CTGCTCACTC	CACCCACCC	ACCTCTGCCC	TGTCCTTGGC	AGCTGAAGGA
22851	CACCTTGACT	TCCAGCTTTT	ACGAGTGAGC	CAAAAACAGA	AGGACAAGTA
22901	CAACTGTGCT	GGCTGTCTGT	ACAAGCTTCA	AAAAGTGTC	CAGAGCCCTC
22951	ACGGTCTCGT	GTCAGATGGT	GTGAGGCTGT	CACGGACATA	GGGATAAACT
23001	TGGTTAGGAC	TCTGGCTTGC	CTTCCCCAGC	TGCCCTCACT	CTGTCTCTGG
23051	CAGCTCTGCA	CCCAGGGACC	ATGTGCTCTC	CACACCCAGG	AGTCTAGGCC
23101	TTGGTAACTA	TGCGCCCCCC	GTCCATCATC	CCCAAGGCTG	CCCAAACCAC
23151	CACTGCTGTC	AGCAAGCACA	TCAGACTCTA	GCCTGGACAG	TGGCCAGGAC
23201	CGTCGAGACC	ACCAGAGCTA	CCTCCCCCGA	GACAGCCCCAC	TAAGGTTCTG
23251	CCTCAGCCTC	CTGAAACATC	ACTGCCCTCA	GAGGCTGCTC	CCTTCCCCCTG
23301	GAGGCTGGCT	AGAAAACCCA	AAGAGGGGGA	TGGGTAGCTG	GCAGAATCAT
23351	CTGGCATCCT	AGTAATAGAT	ACCAGTTATT	CTGCACAAAA	CTTTTGGGAA
23401	TTCTCTTTTG	CACCCAGAGA	CTCAGAGGGG	AAGAGGGTGC	TAGTACCAAC
23451	ACAGGGAAAA	CGGATGGGAC	CTGGGCCCCG	ACAGTCCCCC	TTGACCCCGA
23501	GGCCCATCAG	GGAATAGCCT	CCCTTTGGTA	AATCTGCCTT	ATCCTTCTTT
23551	ACCTGGCAAA	GAGCAATCTA	TGTTAACTCT	TCCTTTATCAG	CCTGTGGCCC
23601	AGAGACACAA	TGGGGTCTTT	CTGTAGGCCA	AGGTGGAAGT	CCTCCAGGGA
23651	TCCGCTACAT	CCCTTAAGTG	CATGCAGATG	TGGAAGGGG	CTGATCCAGA
23701	TTGGGTCTTC	CTGCACAGGA	AGACTCTTTA	ACACCCCTAG	GACCTCAGGC
23751	CATCTTCTCC	TATGAAGATG	AAAATAGGGG	TTAAGTTTTT	CATATGTACA
23801	AGGAGGTATT	GAGAGGAACC	CTACTGTGGG	CTTGAAAAAT	AATAGGTTCC
23851	ATGTGTAAAG	GTTTTGTAAA	ATTTAGTTGA	AAATGCACAG	AAAATCTTCT
23901	GGCCTCTCAT	CACCTGCTTT	CTCAAGCTTC	TTGAGCTTAA	CAACCCCTTC
23951	CCTAACAGGT	TGGGCTGGCC	CAGCCTAGGA	AAACATCCCC	ATTTCTAACT
24001	TCAGCCAGAC	CTGCGTTGTG	TGTCGTGTGT	TTGAGTGAGC	TGGTCAGCTA
24051	ACAAGTCTTC	TTAGAGTTAA	AGGAGGGGGT	GCTGGCCAA	AGCCAACACA
24101	TTCTTGGCCC	AGGAGCAATT	CTTTTCTGGT	AATTCATTAT	GCCATCTGGC
24151	TGCCAATTGA	ACTCAAAACT	TGGAAGCGGA	AGGACAATTG	TATCTGGGAT
24201	TCACCGTGCA	CAGCACCCGA	AGTGCCAAAT	TCCAGGAGGA	CAAGAGCCTT
24251	AGCCAATGAC	AACTCACTCT	CCCCTACTCC	ACCTCCTTCC	AAGTCCAGCT
24301	CAGGCCAGAG	AGGTGGGAGA	AGGTCACAGA	GCCTCAGGAA	TTTCCAAGTC
24351	AGAGTCCCCCT	TTGAACCAAG	TATCTAGATC	CCCTGAGGAC	TTGATGAAGT
24401	GATCCTTAAAC	CCCCAAGTAA	TCATTAACCC	CCAGACCAGC	CTCAGAAGTC
24451	AAGGAGATTG	TTGACCCAGT	GACCTGGAGT	TGAGGCTCAG	GGAGAGATCT
24501	GCCACATGTC	TGAGGGTTGC	AGAGCC		

FEATURES:

Start: 1997
Exon: 1997-2121
Intron: 2122-4732
Exon: 4733-4872
Intron: 4873-5004
Exon: 5005-5115
Intron: 5116-5781
Exon: 5782-5957
Intron: 5958-7770
Exon: 7771-7935
Intron: 7936-8470

Exon: 8471-8623
 Intron: 8624-8917
 Exon: 8918-9000
 Intron: 9001-9777
 Exon: 9778-9925
 Intron: 9926-10221
 Exon: 10222-10335
 Intron: 10336-10539
 Exon: 10540-10617
 Intron: 10618-11197
 Exon: 11198-11293
 Intron: 11294-13338
 Exon: 13339-13445
 Intron: 13446-14214
 Exon: 14215-14284
 Intron: 14285-14400
 Exon: 14401-14493
 Intron: 14494-15980
 Exon: 15981-16262
 Intron: 16263-17597
 Exon: 17598-17652
 Intron: 17653-18842
 Exon: 18843-18988
 Intron: 18989-20477
 Exon: 20478-20549
 Intron: 20550-22478
 Exon: 22479-22523
 Stop: 22524

CHROMOSOME MAP POSITION:

Chromosome 1

ALLELIC VARIANTS (SNPs):

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
48	C	G	Beyond ORF(5')			
132	G	A	Beyond ORF(5')			
724	A	C	Beyond ORF(5')			
1558	C	G	Beyond ORF(5')			
1577	A	G	Beyond ORF(5')			
2487	C	A	Intron			
2634	T	C	Intron			
4352	A	G	Intron			
5157	A	C	Intron			
5658	A	T	Intron			
5945	T	C	Exon	180	T	T
6281	C	T	Intron			
6452	G	C	Intron			
6610	T	G	Intron			
7247	T	C	Intron			
7360	A	G	Intron			
7644	A	T	Intron			
8127	A	C	Intron			
8317	G	A	Intron			
9079	G	A	Intron			
9537	G	T	Intron			
12302	C	G	Intron			
12354	C	T	Intron			
12487	C	T	Intron			
13198	-	A	Intron			
13257	A	G	Intron			
14541	G	A	Intron			

14545	A	G	Intron			
15041	C	A	Intron			
15053	A	C	Intron			
15065	A	G	Intron			
15108	A	C	Intron			
16274	-	G	Intron			
17424	C	T	Intron			
17627	G	A	Exon	657	V	V
18427	T	C	Intron			
18813	C	G	Intron			
19035	T	C	Intron			
19182	T	C	Intron			
19508	-	G C	Intron			
19571	T	G C	Intron			
20147	T	G	Intron			
20180	G	A	Intron			
20584	A	T	Intron			
20717	T	C	Intron			
20894	A	G	Intron			
21787	-	A C	Intron			
22264	T	C	Intron			
22338	-	C A	Intron			
23363	T	C	Beyond ORF (3')			
23688	G	A	Beyond ORF (3')			
24210	A	C	Beyond ORF (3')			

Context:

DNA

Position

48	CTGGGTTTCCTATGTGGGGAGGTCATGCTCCCCACTCATTGAGCCCCC [C,G] CAGGCAAACCACCTGGACAGCCAGACCCATGCAGACTCTGGAGCAGGTGGAGAGGAAGAG TGAGACCACCCCGCCTCACGGGCGGTGAAGGGCCGGCAGCCTCTGAATAGTCTCTGCTAG GAGGTAGAAAGCACCCCTCCCATCTTAATCATAGTAATCATCGCCACTACCATTACTGGG TGCCTATAAAAGGCCAGCCTCTTCATACACATGATCTCACTGAATCCTCATAGCATCTGC CTGCGACTGTTATTATCCCATTACAGATGAAGAACTGAATCTTTGAACCCAGGTCAT
132	CTGGGTTTCCTATGTGGGGAGGTCATGCTCCCCACTCATTGAGCCCCCCCAGGCAAACCAC CTGGACAGCCAGACCCATGCAGACTCTGGAGCAGGTGGAGAGGAAGAGTGAGACCACCCC GCCTCACGGGC [G,A] GTGAAGGGCCGGCAGCCTCTGAATAGTCTCTGCTAGGAGGTAGAAAGCACCCCTCCCATCT TAATCATAGTAATCATCGCCACTACCATTACTGGGTGCCTATAAAAGGCCAGCCTCTTC ATACACATGATCTCACTGAATCCTCATAGCATCTGCCTGCGACTGTTATTATCCCATT ACAGATGAAGAACTGAATCTTTGAACCCAGGTCATCTGGCTCTCAAACCTTGTGCTGTTT TCCCTAAGCCACCCGGTCTCTCATTTCTCCCACTGAAATGTCTCACATGCCATTGCCCTT
724	ATTGCCCTTACTCATTTCTGCCCATGTCTCCTCCAAAACACCATTTATCAATTCGCTCAA CAAGTATGTGTTGAGTACACACTAAGGGCCAGGCGAGGGGCTGGGCACAGGCGCTGGGGG TAGGTTCAATCTCCACCTTCGCTTCTGCTGGGTATCACCTGTGGGGTCTTGCCGGGCAT CCCACCTCACCTGTAGTTCAAGTGGACCTTGGGATCCCAAGACCAATGAATGGAATGC ACCAGCCAGCCTTCACCAACTTGAGCACAATCTTATTATAATAGAACTCACATTTGC [A,C] TCACACTTTACATTTTACACAACCCCTTCTTATCCATTAACCTCATTTGATCTTCACAACA ACCCTGTGAGATATGTCTGTACTCCCACTTTAGTGATACAGAATCTGAGGTTTGAAAAG TAATGCTGACCATTTCTGCCTCATTAATAAAAGCAGGATTAACCCAGGCTCCTGGACCCCT CCACAAAAGGCATTAAGCAACCTGCTCCCTCTGACAACCTCCCTGTACCCAGGCTCT CCTCTGGGAAGTTGGGGGCATCTCTAGCCCCAAGTAGTTACTCATTTTCAACCCATCT
1558	TCAGCTCTGCCCATCTCAGCTCCTGGAACGTGAGCCAGGTTGCGCAAAAAGTGAGGAGGA GAGGAGCGGCAGTACACAAGGTGGGGGAAAGATTAGGCACAGGAAGCCGTGGGAGAGAG AGCCGGCAGGTGGACCATCTTGTTTCCCCACACACACCATTTGTCCCCTGGGAAACCTG

FIGURE 3, page 10 of 20

TTGGTGAAGTTCTAGATGTCTTATCCAAGAAGGGTCCTCTTGAGGTCTCTCAGCTATCC
CCCTGCCCTCTAGGCAAGCTGTTTTCTGTTTCTTCCAAGCTGACTGGCTGAATGGTAGGAG
[C, G]
CTTTCTGCCAGGAACTAAGGTCTGGGAAGGGAGTATGGCTTGTGGGGACACCAGGGGT
CAGGGGAGGGGAGGGTCCACCTGCTGAATCAAGTGGGGCCTCTGCCCTCGTGATTCCCC
TTTGCTTGGTGCTCAGTGGGGGTGATGGTGACGCCACAGGTGTGGAGTGCCAGCCACGTG
CTGAGCGCCAAGCAAAACAGCCAGGGTGAGTCTATGCATCATCAGTGCCTGGGAAGGAAG
GCCACTGCGAGCAGGGAGTCTGACGGAAAACTTGACAGAGGGAAGGGAGGCACCTTGCT

1577 CTCTGGAACGTGAGCCAGGTTGCGCAAAAAGTGAAGGAGAGAGGAGCGGCAGTACACAA
GGGTGGGGGAAAGATTAGGCACAGGAAGCCGTGGGAGAGAGAGCCGGCAGGTGGACCATC
CTGGTTTCCCCACACACACCATTGTCCCCCTGGGAAACCTGTTGGTGAAGTTCTAGATGT
CTTATCCAAGAAGGGTCTCTTGAGGTCTCTCAGCTATCCCCCTGCCTCTAGGCAAGCT
GTTTTCTGTTTCTTCCAAGCTGACTGGCTGAATGGTAGGAGCCTTTCTGCCAGGGAACT
[A, G]
AGGTCTGGGAAGGGAGTATGGCTTGTGGGGACACCAGGGGTGAGGGGAGGGGAGGGTCCA
CCTGCTGAATCAAGTGGGGCCTCCTGCCCTCGTGATTCCCCCTTGCCTGGTGTCTCAGTGG
GGGTGATGGTGACGCCACAGGTGTGGAGTGCCAGCCACGTGCTGAGCGCCAAGCAAAACA
GCCAGGGTGAGTCTATGCATCATCAGTGCCTGGGAAGGAAGGCCACTGCGAGCAGGGAGT
CTGACGGAAAACTTGACAGAGGGAAGGGAGGCACCTTGCTTTATCGGGGCGGGGAAGGC

2487 ACACGGCTTCTGCACTGGTATCCCTAAGATGGGGTTAAGGGAAGCCCTGGGGAAGTGAGG
TTCTGAATGATGAATTAAGATCCTACAACCTCATCTGTACTGAGACCCCCAGGGAGGAT
GGGAGCAGGAGCAAGAACCATCCAGAAGGGTTATATGGCATTCCCAAACCCCTGCATGG
CATCTCCATATTCTCAATTACCCGGGTCTCTCTGGGTTTGTAAAGGCATGGTAGATGA
GCATCTACGTTATGGAGGGGTGGGGAGCATCAGAGCCCTTACTCCATGCCCTGTTCCCTC
[C, A]
TTACAAAAATACCTGAAGTTACCATCACCCAGGTTCTTTGTCCTTTCCCTCCCGGATG
TTCCTTCCTCCACTTGGTCCAGAGAATGCCAAAAGGAGGCCCTAAATTTCTGAACTTTCC
TGAGGGGACCTACAGGGTGATAGTCTACCAGCGCCAGGGTCTTTCCACTCTCATCTCC
CTGGAAATGCGATGGTGGGTATGAAACCTTGTCCCTAAGTAGGCGCTACACAAGGTGATC
CATACCCACACCCAGGAGGCTGGGGTGTGCGGGTGTACCCTCCCCATTCCCAGACTCCT

2634 AGGGTTATATGGCATTCCCAAACCCCTGCATGGCATCTCCCATATTCTCAATTACCCGG
GTCTCTCTGGGTTTGTAAAGGCATGGTAGATGAGCATCTACGTTATGGAGGGGTGGGGAG
CATCAGAGCCCTTACTCCATGCCCTGTTCCCTCCTTACAAAAATACCTGAAGTTACCAT
CACCCAGGTTCCTTTGTCCTTTCCCTCCCGGATGTTCCCTTCCTCCACTTGGTCCAGAGAA
TGCCAAAAGGAGGCCCTAAATTTCTGAACTTTCTGAGGGGACCTACCAGGGTGATGCC
[T, C]
ACCAGCGCCAGGGTCTTTCCACTCTCATCTCCCTGGAAATGCGATGGTGGGTATGAAAC
CTTGTCCCTAAGTAGGCGCTACACAAGGTGATCCATACCCACACCCAGGAGGTGGGGC
TGCGGGTGTACCCCTCCCCATTCCCAGACTCCTGGCAGACCTCCTCTGGCCAGCTATAG
GCCAACTCACTCTCCCTCACTCCCTTGGGGAACGGCTGATTCACTTACCTGGATTGAGG
TCACTGGCAATGGCTGAAGTGGAGACGCAGGTGGAAGTGGTTAGGCCGGGGGAATCACC

4352 ATTGGAGTTACCACACATAAAGGATAGTGAGTCAGCAGAGTGCACCCTGCAGGAACAATA
GAGCCTTCCTTTTCAAGGAAGTTCTAAGAAAAATGGCAGCAGGCAGGCCCACTCGGGTG
TATTCATCTATTCAATTTATTCAACAAATATTTACTAAGTGCCCTGTGCAAGGCTCGAGG
TGTACAAAGATGAACAGGAGAGCTAGACTTCTTGCCATGCGTGGTGGGGTTTGCTGCCTA
GTGGGAGAGACAGACAAAAAGCAAGGAATGCACACACAGGATGCACACACAGCGGCAGGA
[A, G]
CCAAGGTGCAGTTACCCAGGCTGGGATCAGACAGACAGGACTCAGAGGAGACTTTCCCA
GAGAAAAGCCATCTGAGCCAAGGGATGGATCTGATACCTCCGAAGGCTGAGCCACCATAA
CACTCATACCTTTAAGCCAAGTCTTATAAACTCCCCAGGTAAGCAGCTGGCAGTCAGAAG
ACCTCCAGCTAATGCCAGGACAAGTTGATGAGCTCTCAAGAAAAAGTTCTGCCTTTTC
TTCTCAATATCCCTGGCACACAGTTCACTGAATTTTGAATGAACCAATGAATGAAATGAG

5157 ATCCAGGTCCCACAAGGTGAAGGGCTCCTTCAGCCAGGCCTGGATTGCCACTCCCCTCA
CCATTCTCTCTCATCCCCACTCCATCCCTCTGTGATCCCCATAAGCTAGTCATGCTGC
TGAGCTTCAGTCTCGTTGTCTCTGACGGCATGGCATTGCTCTGCTGGCCAACCTTCCT
GCAGTCAATGGCTCTACTCCTCCTTCTTCCCCCTCCTGACCTACTTCTTCTGGGGGGT
GTTACCCAGATGGTGCCAGGTAAGGCCTCTCCCTCTGGGCAGGCAGGATGACCCAGACC
[A, C]

FIGURE 3, page 11 of 20

CAAGGATGGGAGGTGTGGCAAAGGGGCCTCGGGAGATTTCCATCTGCATTCTCCTGGAG
TTGTTCTCTGGTCAGTCCTAGGGGAATGGTCACTGTGAATGTCAATTCAGGTCCTCGGTG
ACCTTGGAGAAACCACTGAGCCTCTTTGAGTTCAGTTAGCATTACCTGTTCCATCTTCCT
CCTAGGAATGAGAGGAAGACTTAGCAGAACAAGATATACCATATGCTATAACATGCTTAA
ACAGATGTGAGAAATCACCATCTAACTCCCTGGTTGGTCCCAGCCGGCCACTACAGGGAC

5658 TTAGCAGAACAAGATATACCATATGCTATAACATGCTTAAACAGATGTGAGAAATCACCA
TCTAACTCCCTGGTTGGTCCCAGCCGGCCACTACAGGGACATTTGGACTTCTCTGGTGC
AAGTGAGATGGAGGAAAGCCTGGTCAACAGGGCTGGTTTCTGGTTCAGGCTCTGCTTATA
TTTCTTATTTCTGAGTTCATTTTCTCACGTGCTCTGTATGACAATATTGACCATTGGGGT
AAAAGCACCTTGAAAAGCATAGATCATGGTTAGAGTGAGTGGTTGTTATTATTGTGTGG
[A, T]
GAAGAGCCTTGGAGGTGCAGGGATCCATCCCCCTGGGGTCGGGAAGCATTCCTGGGCCCC
TTTCTGGTTTCCATCGGTGTGGTTCAAACCTCTGATTTTGTGGCTGGGTGGGGCACCA
CAGGTACCTTTGCCGTTATCAGCATCCTGGTGGGTAACATCTGTCTGCAGCTGGCCCCAG
AGTCGAAATTCAGGTCTTCAACAATGCCACCAATGAGAGCTATGTGGACACAGCAGCCA
TGGAGGCTGAGAGGCTGCACGTGTGAGCTACGCTAGCCTGCCTCACTGCCATCATCCAGG

5945 ATTATTGTGTTGGAGAAGAGCCTTGGAGGTGCAGGGATCCATCCCCCTGGGGTCGGGAAG
CATTCCTGGGCCCCCTTCTGGTTTCCATCGGTGTGGTTCAAACCTCTGATTTTGTGGC
TGGGTGGGGCACCAAGGTACCTTTGCCGTTATCAGCATCCTGGTGGGTAACATCTGTCT
GCAGCTGGCCCCAGAGTCGAAATTCAGGTCTTCAACAATGCCACCAATGAGAGCTATGT
GGACACAGCAGCCATGGAGGCTGAGAGGCTGCACGTGTGAGCTACGCTAGCCTGCCTCAC
[T, C]
GCCATCATCCAGGTGAGGGGGCAGCCCCAACCCCTGCTAGAAGGGCATCAGACCACCCTG
CCCCCTCCCTCAAAGCCTTAGCTTTGATGCTAAATCTGATTTAGGGGGCTGGGTGTGGAGG
CTCATGCCTGTAATCCAGCACTTTGGGAGGCTGAGGAGGGTGGATCACTTGAGGTGAGG
AGTTTGAGACCACCTTGACCAACGTGATGAAACCCCATCTCTACCAAAAATACAAAATA
ATCCAGGCTTGGTAGTATGCGCCTGTAGTCCCACCTACTCAGGAGGCTGAGGCAGGAGAA

6281 GCTAGAAGGGCATCAGACCACCCTGCCCTCCCTCAAAGCCTTAGCTTTGATGCTAAATC
TGATTTAGGGGGCTGGGTGTGGAGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGA
GGAGGGTGGATCACTTGAGGTGAGGAGTTTGGAGCACCTTGACCAACGTGATGAAACCC
CATCTCTACCAAAAATACAAAATAATCCAGGCTTGGTAGTATGCGCCTGTAGTCCCACC
TACTCAGGAGGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGGTTGCAGTGAGC
[C, T]
GAGATCGCGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCCGTCTCAAAAAAAA
AAAAAAAAAAAAAAAAAAAAACCAAGTTAGGGCTCACCTCCTCCCTCCTCCCCATCCCAGG
GCTAAAGTGAACCTTGAAAATTAACAGTATCTCCTCATCTGCATGTAGCAGGACCATA
AAAAAACACAGCTGTACCTGGTTAAACTGTCTGAGCTTTAAACCTGTAAAAGACTCAC
AGCCTCTCTCCATTATCCCGTGGAGAAACCAACTCTCTGCCAGCATAGTCTTGCACT

6452 ATGAAACCCCATCTCTACCAAAAATACAAAATAATCCAGGCTTGGTAGTATGCGCCTGT
AGTCCCACCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGGTT
GCAGTGAGCTGAGATCGCGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCCGTCT
CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCAAGTTAGGGCTCACCTCCTCCCTCCTCC
CCATCCCAGGCTAAAGTGAACCTTGAAAATTAACAGTATCTCCTCATCTGCATGTAGCA
[G, C]
GACCATACAAAAAACACAGCTGTACCTGGTTAAACTGTCTGAGCTTTAAACCTGTAA
AAGACTCACAGCCTCTCTCCATTATCCCGTGGAGAAACCAACTCTCTGCCAGCATAGTC
TTGCAGACTGCTAATTTTCTCTAACATCCCTCACTCCGCTCCAGCCTCCTCTGCTCCAAG
CCACAGCAGCAGTTGCACAACATAAATTGAGCTTCTGCAAATGGTTGCAAAGGATTCTGC
TAGGTTTTATGAAGGAAGCACAAATGACAGAATGCAAGAGCAAAACACAGTCCCAGAG

6610 GTGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCAAGTTA
GGGCTCACCTCCTCCCTCCTCCCATCCAGGGCTAAAGTGAACCTTGAAAATTAACAGT
ATCTCCTCATCTGCATGTAGCAGGACCATACAAAAAACAACAGCTGTACCTGGTTAAAC
TGTCCTGAGCTTTAAACCTGTAAAAGACTCACAGCCTCTCTCCATTATCCCGTGGAGAAA
CCCAACTCTCTGCCAGCATAGTCTTGCACTGCTAATTTTCTCTAACATCCCTCACTCC
[T, G]
CTCCAGCCTCCTCTGCTCCAAGCCACAGCAGCAGTTGCACAACATAAATTGAGCTTCTGC
AAATGGTTGCAAAGGATTCTGCTAGGTTTTATGAAGGAAGCACAAATGACAGAATGCA
AGAGCAAAACAGTCCCAGAGAGCGCCTTTTCATTCACTCATTTCGGTTTTGTGCC

FIGURE 3, page 12 of 20

AAGAACTAGGCTAAACCTGGGATACAAAGATAAGTAAGAAAGAGGTCCAATTACAAAGT
TGCTCACAGCCCAGCAGAGGAAGGAGCCATGTCAACAGATAAATTTGTATGCAGTGAGAT

7247 GACACAGAGCAGAGTCACGGAGGACCTCAAAGAGGAGGTGACACTCCACCTCTCTTAAAG
GATGAGAACTTAACCAGGAACAAGGTATACAGAGGATGGTCCAGGCAGAAGGGAACAGTG
CCTAAAAACACTGAGGCCTGAGAGAGTGTGATCTGCGCAGGCAAAGTAAGGGGCTTGGTG
TGGCTGGAGGGTAGAGGGCCCAGAAGAGGATGGAAAAGTAGGCAGGAGCCAGACAATGAG
ATCTGGGGTCTGTTCTCTGACAGCGACTTTGGGTCTGATTGGCAGTTTATAAGGATCGTT
[T, C]
GGGCTACACAATGATGAGTGGGAGGTGGATTAGAATCAAGGCAGGGGACCTGTTGGGAGA
CTCTGCAGAGGGCCAGGCAGGAATAATGCAGGCGAAGACCAGGTAGAGAAAGAGATGGGG
CTGGACTTGAAAAGAATGTTTTACCAGGAGCTTGGTGATAGACTGGATGTGGGAGGTAAG
GGAGGATGACTCTCAAGTTTTTGGTTGGGCAACCAGGTTAATGATGGTGTATTTACTGA
GAGAGAAAACACTGGGGGAGGACTAGACTTATTTTACAGATAAGCCAAAGCCAGAGAGGT

7360 AACAGTGCCTAAAAACACTGAGGCCTGAGAGAGTGTGATCTGCGCAGGCAAAGTAAGGGG
CTTGGTGTGGCTGGAGGGTAGAGGGCCCAGAAGAGGATGGAAAAGTAGGCAGGAGCCAGA
CAATGAGATCTGGGGTCTGTTCTCTGACAGCGACTTTGGGTCTGATTGGCAGTTTATAAG
GATCGTTTGGGCTACACAATGATGAGTGGGAGGTGGATTAGAATCAAGGCAGGGGACCTG
TTGGGAGACTCTGCAGAGGGCCAGGCAGGAATAATGCAGGCGAAGACCAGGTAGAGAAAG
[A, G]
GATGGGGCTGGACTTGAAAAGAATGTTTTACCAGGAGCTTGGTGATAGACTGGATGTGGG
AGGTAAGGGAGGATGACTCTCAAGTTTTTGGTTGGGCAACCAGGTTAATGATGGTGTCTAT
TTACTGAGAGAGAAAACACTGGGGGAGGACTAGACTTATTTTACAGATAAGCCAAAGCCA
GAGAGGTGATGTGACAGAAAGGCCCATGCTCTAAAGGAGCTGAAGGTCTGATGGCAGCCA
TGTAGAGCACAGTGAAGGGCAGGTGAAGGTCACAGATGGTCCAATTCCCTCAAGCTACTG

7644 GACCAGGTAGAGAAAGAGATGGGGCTGGACTTGAAAAGAATGTTTTACCAGGAGCTTGGT
GATAGACTGGATGTGGGAGGTAAGGGAGGATGACTCTCAAGTTTTTGGTTGGGCAACCAG
GTTAATGATGGTGTCTTTACTGAGAGAGAAAACACTGGGGGAGGACTAGACTTATTTTA
CAGATAAGCCAAAGCCAGAGGATGATGTGACAGAAAGGCCCATGCTCTAAAGGAGCTGA
AGGTCTGATGGCAGCCATGTAGAGCACAGTGAAGGGCAGGTGAAGGTCACAGATGGTCCA
[A, T]
TTCCCTCAAGCTACTGCTACGCTAGGACTGCACGGAGCTCCAGACCTGCGTGTGTGGG
GCGGGTCGTTGGAAGTCTGAACACATTGGTCTTCCGCCACCAACCACCTTTTCCTCC
TCTCAGATGGGTCTGGGCTTCATGCAGTTTGGCTTGTGGCCATCTACCTCTCCGAGTCC
TTCATCCGGGGCTTCATGACGGCGCGGCTGCAGATCCTGATTTTCGGTGTCTCAAGTAC
ATCTTCGGACTGACCATCCCCTCCTACACAGGCCAGGGTCCATCGTCTTTGTGAGTCTG

8127 CATCCGGGGCTTCATGACGGCCGCCGGCTGCAGATCCTGATTTCCGGTGTCTCAAGTACAT
CTTCGGACTGACCATCCCCTCCTACACAGGCCAGGGTCCATCGTCTTTGTGAGTCTGGG
GATGCACCCCTGCCATTGGAGCAAGGCTCCAGCAGACACATGAGGAGGATGTACTGTTTT
AAGATGTCGTGAGCTCCTCATTGCAAGGGCTGGCTTAGCTGTTGTTTACAGAGAGGATTCTG
AGGGGGTTTCTGTCTTGGGAGGGTCAAAGTCATGACTCACAGAGGTTCTTGGTAGTTAAT
[A, C]
CCTGCAGAAAAGAGCTGTACATTCTCCGCCAGTTCCCCATTCTAGTGCCTCAACCCCTCC
CTGCCTGGAAAAGTCTGCCTTATGTCTAATCTCCATCCCTCCTCCTTCAGCCCAAACCTCT
TCTAAAGAAAAGAAAGCATTCTTTTCTAGCACAAAGTTCCCCATGTGCCTTTTGGGAAA
GGGCGGTGGGCGACGGGACAGGGTTCTGATCAGGGTTTAAATTCTGTCTTGGTGTGCCT
CCATTAGCTTTGATGGCATCCCTTCCCTGGGTGAGACACCCAAAGGTGGGGTATTATGGG

8317 GAGCTCCTCATTGCAAGGGCTGGCTTAGCTGTTGTTTACAGAGGATTCTGAGGGGGTTTC
TGTCTTGGGAGGGTCAAAGTCATGACTCACAGAGGTCTTGGTAGTTAATACCTGCAGAA
AAGAGCTGTACATTCTCCGCCAGTTCCCCATTCTAGTGCCTCAACCCCTCCCTGCCTGGA
AAGTCTGCCTTATGTCTAATCTCCATCCCTCCTCCTTCAGCCCAAACCTCTTCTAAAGAA
AAAGAAAGCATTCTTTTCTAGCACAAAGTTCCCCATGTGCCTTTTGGGAAAGGGCGGTGG
[G, A]
CGACGGGACAGGGTTCTGATCAGGGTTTAAATTCTGTCTTGGTGTGCCTCCATTAGCTT
TGATGGCATCCCTTCCCTGGGTGAGACACCCAAAGGTGGGGTATTATGGGAAGAAGGGGT
GGGAGCCTGTGAGCATGATGCTCTTTCCCCCAGACCTTCATTGACATTTGCAAAAACCTC
CCCCACACCAACATCGCCTCGCTCATCTTCGCTCTCATCAGCGGTGCCTTCCCTGGTGTG
GTGAAGGAGCTCAATGCTCGCTACATGCACAAGATTGCTTCCCATCCCTACAGAGATG

FIGURE 3, page 13 of 20

9079 TTCTACTGCTCTAATAATTCCCCTAAGGAGGCAGGGGAGTGGGATTGAGGTCCCCAGA
GAAAAGGGGAGACTTGAGAGAGACGCCTGCCCTGGCCCCACCTTAGGGCCAATCCCCATT
TCCACTCTGGGGTTTGCAGGTGGTGGTGGCAACAGCTATCTCGGGGGCTGTAAGATGCC
CAAAAAGTATCACATGCAGATCGTGGGAGAAATCCAACGCGGGTGAGTCCAGGTGGCCCA
GAAGCCTGGCCACCCGCACCTCATGCCCCACTAAGGCCTGAGCTCGGAGAGGGAGACAA
[G,A]
ATGAACTCTATGAAAGTGCAGTCGAAACTGTATGACACTGACCATGTATGAATTATTACT
ATTACCGTTTCTGAGAAGGGCCGCACAACAGCCAATGTAGGCTATTTTATGAGAAATG
AGTCTTAACTGCCACACTCCCCTTATAAATCTCATTCAACTGATGCTGTTAAACAAAGCC
TCTCTGAACAGCCGCTTGCTGGCTCTTTGCCTTGCTCTAATGCATTGGTTCTTTGTCCAT
GTAGAAAGGGAATATTAGGTTCAACCAGATTTCATGAAGCATCCACTCTGTGCCAGGCAC

9537 AACTGATGCTGTTAAACAAAGCCTCTCTGAACAGCCGCTTGCTGGCTCTTTGCCTTGCTC
TAATGCATTGGTTCTTTGTCCATGTAGAAAGGGAATATTAGGTTCAACCAGATTTCATGA
AGCATCCACTCTGTGCCAGGCACCATGCTGGGCCCTGGGAGGAGAGGGGTGACGCTTGTC
CTGCAGGGTTGGAACAGGCAAGGGAGGGAAGACCACATAGCACCAAGGTCTAGGGGTCT
GTGGACTCGTGAGCATACAGGGTTCAGAATCTGGGAGTTAACAAACGAGGCCCTACCACA
[G,T]
ACTGGCCCCGGGACCTTGGGCAAGTTAGGTTCTCTCAGCCTCAGTTTCTCTCTTTGTAAA
ACAGGAGTGATGGTCCCTACCTATGGGGTGGTCTGAGGATTGAGACTGGATGGGATAA
CTTAGGCAAAGATCCCGGCACACCATGGGGGCCCTGGCTGGTCCCTGTGGGCTGGTGAAGG
ACTTGGCTGCCCTCCCCACTCACACCTTGGGTTCTGCCTCCTTCTGGCTCCTCGGCAG
GTTCCCCACCCCGGTGTCGCCTGTGGTCTCACAGTGGAAGGACATGATAGGCACAGCCTT

12302 AGCCCCACCATAACCTATGGGAGAGGATTTACTAACTTTCTTAACGGTGATGAAACCAA
GGCTCAGAATGGTTAAGTAAATTGTCAAAGGCCACAGAGGTAGGGAGTGGTAGAGTCTGG
ATTAAACTTCCAAGTCTTGACTCCAGACCTCTAGGCTGTACTGTCTCATAGGGAAGGCA
GTCTCACCACCTAGGGCAGAGAAGAAAATCCTTAAAGCCAGAGAAGTGAAGTGGCTCATC
TGTGGTCACCCAGAGAGACAGTGATGAGGACAGGGAGAAAAATTATACCTCAGTTCACG
[C,G]
CCAAGGATCTGCTTTGACCATAACCCAACAAGCCCCGCTATGGTGGTATTTCTTAGGT
TCATATGGCGGCTTTTGTTCATTGTATCTTACAGCAATTCTCTACAGGAATCTGGGC
AGATTTATTTCTTTAGAGGAATTTCCAGGTCTTAAATCTATAGGGGGCAACTATCAAA
ACTTCACCCAATGTTGCCCCCTACCCACACACAAAACAGGCCCCAGCCGATCAGAAAG
CACTGCTGAGCTCCTGTGAGGGCCACGCAGCTCGCTGTGAGACAGAGAGAGGGAATCA

12354 GAAACCAAGGCTCAGAATGGTTAAGTAAATTGTCAAAGGCCACAGAGGTAGGGAGTGGTA
GAGTCTGGATTAAACTCCAAGTCTTGACTCCAGACCTCTAGGCTGTACTGTCTCATAG
GGAAGGCAGTCTCACCACCTAGGGCAGAGAAGAAAATCCTTAAAGCCAGAGAAGTGAAGT
GGCTCATCTGTGGTCACCAGAGAGACAGTGATGAGGACAGGGAGAAAAATTATACCTCA
GTTCCCAGCCCAAGGATCTGCTTTGACCATAACCCAACAAGCCCCGCTATGGTGGTATT
[C,T]
CCTTAGGTTTCATATGGCGGCTTTTGTTCATTGTATCTTACAGCAATTCTCTACAGGA
ATCTGGGCAGATTTATTTCTTTAGAGGAATTTCCAGGTCTTAAATCTATAGGGGGCAA
CTATCAAACTTCACCAATGTTGCCCCCTACCCACACACAAAACAGGCCCCAGCCGA
TCAGAAAGCACTGCTGAGCTCCTGTGAGGGCCACGCAGCTCGCTGTGAGACAGAGAGAG
GGAATCACATTTATTGATCACCTACTGAGCATCCATCACTAGGCTAGGACCGTCACATT

12487 ACCCACCTAGGGCAGAGAAGAAAATCCTTAAAGCCAGAGAAGTGAAGTGGCTCATCTGTGG
TCACCCAGAGAGACAGTGATGAGGACAGGGAGAAAAATTATACCTCAGTTCAGGCCCAA
GGATCTGCTTTGACCATAACCCAACAAGCCCCGCTATGGTGGTATTTCTTAGGTTTCAT
ATGGCGGCTTTTGTTCATTGTATCTTACAGCAATTCTCTACAGGAATCTGGGCAGAT
TTATTTCTTTAGAGGAATTTCCAGGTCTTAAATCTATAGGGGGCAACTATCAAACTT
[C,T]
ACCCAATGTTGCCCCCTACCCACACACAAAACAGGCCCCAGCCGATCAGAAAGCACTG
CTGAGCTCCTGTGAGGGCCACGCAGCTCGCTGTGAGACAGAGAGAGGGAATCACATTT
ATTGATCACCTACTGAGCATCCATCACTAGGCTAGGACCGTCACATTCTTAACTTTGA
ATCCTTTTCATGAGGTAGGCATTATTATTCTCCTTTTGTTCACATAGCCATTAAAGAACA
AAATTTGGGGCTGGGTGTGCTGACTCACACCTGTGATCTAGCACTTAGGGGGCTGAGGC

13198 CTAACCTATTAGGAAGGTTAGGCGGGAGCACAACTTGGGTTCCAGGTTTGAAGGCTCCAG
TGAGCTGATCTTGCCACTGCACTACAGCCTGAGCAACAGAGCAAGACCCTGTGACTCCAA
AAACAAACAAACACATTTTGAACCCAAACAGATCTGACCCAAGATGCATGCTCTTA

FIGURE 3, page 14 of 20

TAGATGCCACCTCCCTGTGTGCTGGGGCTTCTACTAAAAACACAGACAAGATCAGGCAAC
CACAGTCAATCTAAGGGAAAGAGGAAAGTGTAAACAAAGCACAAATACATAAAATATTGC
[-, A]
AAAATGCTATTTAAAGAAAAAAGAGAAGAGAGGCTCTGAGGTTGTACTAACAGAGAAT
GGCCTTGGCTAATCCAGGAAGACTTCTGAAAGAGGTTGTTTTTCCCCAGGTCTGCTTT
TGACATCTCTTTTTACAGTGCATCTGGGTAGTGAGCTTCCCTCCTCCTTCTTCTCA
GCCTGCCCTATGGTGTGGCAGTGGGTGTGCGCTTCTCCGTCCGTGGTGGTCTTCCAGA
CTCAGTTGTAAGTGATAGCTTCCGCCCTCCTAGGCCACAGTCGGTTCCTGGGCCAGCC

13257 GTGAGCTGATCTTGCCACTGCACTACAGCCTGAGCAACAGAGCAAGACCCTGTGACTCCA
AAAACAAACAAACAAACACATTTTGAACCCAAACAGATCTGACCCAAGATGCATGCTCTT
ATAGATGCCACCTCCCTGTGTGCTGGGGCTTCTACTAAAAACACAGACAAGATCAGGCAA
CCACAGTCAATCTAAGGGAAAGAGGAAAGTGTAAACAAAGCACAAATACATAAAATATTG
CAAAAATGCTATTTAAAGAAAAAAGAGAAGAGAGGCTCTGAGGTTGTACTAACAGAGA
[A, G]
TGGCCTTGGCTAATCCAGGAAGACTTCTGAAAGAGGTTGTTTTTCCCCAGGTCTGCTT
TTGACATCTCTTTTTACAGTGCATCTGGGTAGTGAGCTTCCCTCCTCCTTCTTCTC
AGCCTGCCCTATGGTGTGGCAGTGGGTGTGCGCTTCTCCGTCCGTGGTGGTCTTCCAG
ACTCAGTTGTAAGTGATAGCTTCCGCCCTCCTAGGCCACAGTCGGTTCCTGGGCCAGC
CCGCAAAGGGCTTCCATGCCACGGCTGGCTTAGTCCACTGTACCTTCACCTCTGGGCC

14541 TCATGGACACTGACATTTATGTGAATCCCAAGACCTATAATAGGGTAGGTAATTCAAGCT
TATGACCTCCTTCTTTTGTCTGTGCACCAACCCCAAGAAGAGGTTGCTTTTTAAAGCCAATA
AAGACATTTCTGCAACTTGAGCTCAGTCTCCCTGTCACAGGCCCAGGATATCCAGGGGAT
TAAAATCATCACGTACTGCTCCCTCTCTACTTTGCCAACTCAGAGATCTTCAAGGCAAAA
GGTCATCGCCAAGGTAAGGCTCAGTCCCTGGCGACCAGAGGCTCTGGACAGAGAGTGGCC
[G, A]
GAAAATGGAAGCAGAAGGGCGGTGGGAGCTGAGAATAGGCCACTCCCATAGAGGGTGGAG
GTCAAGATTGCTGTTGGCTCTCTCCCTGCAGACAGGCATGGACCCCAAGAAAGTATTACT
AGCCAAGCAAAAATACCTCAAGAAGCAGGAGAAGCGGAGAATGAGGCCACACAACAGAG
GAGGTCTCTATTATGAAAACCAAGGTGAATGAAGGCCAGAAGCAGCCCCGTGCCCTGCT
CTCCTGCCCATTCTGATACTGCCCCCTGTTACTCATGGTACCCTGGGGGCCCGCTTCCC

14545 GGACACTGACATTTATGTGAATCCCAAGACCTATAATAGGGTAGGTAATTCAAGCTTATG
ACCTCCTTCTTTTGTCTGTGCACCAACCCCAAGAAGAGGTTGCTTTTTAAAGCCAATAAAGA
CATTTCTGCAACTTGAGCTCAGTCTCCCTGTCACAGGCCCAGGATATCCAGGGGATTAAA
ATCATCAGCTACTGCTCCCTCTCTACTTTGCCAACTCAGAGATCTTCAAGGCAAAAGGTC
ATCGCCAAGGTAAGGCTCAGTCCCTGGCGACCAGAGGCTCTGGACAGAGAGTGGCCGGAA
[A, G]
ATGGAAGCAGAAGGGCGGTGGGAGCTGAGAATAGGCCACTCCCATAGAGGGTGGAGGTCA
AGATTGCTGTTGGCTCTCTCCCTGCAGACAGGCATGGACCCCAAGAAAGTATTACTAGCC
AAGCAAAAATACCTCAAGAAGCAGGAGAAGCGGAGAATGAGGCCACACAACAGAGGAGG
TCTCTATTTCATGAAAACCAAGGTGAATGAAGGCCAGAAGCAGCCCCGTGCCCTGCTCTCC
TGCCCATTTCTGATACTGCCCCCTGTTACTCATGGTACCCTGGGGGCCCGCTTCCCACCC

15041 ACCAAGGTGAATGAAGGCCAGAAGCAGCCCCGTGCCCTGCTCTCCTGCCATTCTGATAC
TGCCCCCTGTTACTCATGGTACCCTGGGGGCCCGCTTCCCACCTGACAGGCAAGACA
GAAAGTCTCTGGGAACACTGCCTGGTGGCCGCTGGGCATTTTTCTTCTTTTTTCTTTT
TCTTTTTAGAGATGGAATTTTGTCTCTGTCAACCAGGCTTGAGTGCAATGGCGTTATCTT
GGCTCACTGCAACCTCCACCTCTGGGGTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGT
[C, A]
GCTGAGATTACAGGTGCCACCACACCCAGCTAATTTTTGTATTTTGTAGTAGATATTGGGT
TTCACCATGTTGGCCAGGCTGGTGTCAAACCTCCTGACCTCAGGTGATCCACCTACCTTAG
CCTTCCAAGTGTCTGGGATTACAAGCCTGAGCCACTGCGCCACGCTGGGCATTTTTCTT
CTTGGATGAGGTGCTACCATCTCCCAGGGAAGCCACTGAACCCCAAGGCCCTTCTCCAT
TTTCTGGCTAAGATAGGACATGGCCATGGACTTTTGAACAACCCAGAGGGGAACAGCA

15053 GAAGGCCAGAAGCAGCCCCGTGCCCTGCTCTCCTGCCATTCTGATACTGCCCCCTGTTA
CTCATGGTACCCTGGGGGCCCGCTTCCCACCTGACAGGCAAGACAGAAAGTCTCTGG
GAACACTGCTGCTGGTGGCCGCTGGGCATTTTTCTTCTTTTTTCTTTTTCTTTTAGAGA
TGGAATTTTGTCTTTGTCAACCAGGCTTGAGTGCAATGGCGTTATCTTGGCTCACTGCAA
CCTCCACCTCTGGGGTTCAAGCGATTCTCCTGCCCTAGCCTCCCAAGTCGCTGAGATTAC
[A, C]

FIGURE 3, page 15 of 20

GGTGCCACCACACCCAGCTAATTTTTGTATTTTTAGTAGATATTGGGTTTCACCATGTTG
GCCAGGCTGGTGTCAAACCTCCTGACCTCAGGTGATCCACCTACCTTAGCCTTCCAAAGTG
CTGGGATTACAAGCCTGAGCCACTGCGCCAGCCTGGGCATTTTCTTCTTGATGAGGT
GCTACCATCTCCAGGGAAGCCACTGAACCCCCAAGGCCCTTCTCCATTTTCTGGCTAAG
ATAGGACATGGCCATGGACTTTTGAACAACCCAGAGGGGGAACAGCAGTGAATTTCTGT

15065 CAGCCCCGTGCCCTGCTCTCCTGCCATTCTGATACTGCCCCCTGTTACTCATGGTACCC
TGGGGGCCCCGCTTCCCACCCCTGACAGGCAAAGACAGAAAGTCTCTGGGAACACTGCCTG
GTGGCCGCTGGGCATTTTCTTCTTTTTTCTTTTTCTTTTTAGAGATGGAATTTTGCT
CTTGTCACCCAGGCTTGAGTGCAATGGCGTTATCTTGGCTCACTGCAACCTCCACCTCTG
GGGTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGTCGCTGAGATTACAGGTGCCACCAC
[A, G]
CCCAGCTAATTTTTGTATTTTTAGTAGATATTGGGTTTCACCATGTTGGCCAGGCTGGTG
TCAAACCTCCTGACCTCAGGTGATCCACCTACCTTAGCCTTCCAAAGTGCTGGGATTACAA
GCCTGAGCCACTGCGCCAGCCTGGGCATTTTCTTCTTGATGAGGTGCTACCATCTCC
CAGGGAAGCCACTGAACCCCCAAGGCCCTTCTCCATTTTCTGGCTAAGATAGGACATGGC
CCATGGACTTTTGAACAACCCAGAGGGGGAACAGCAGTGAATTTCTTGGGGAACCCAGGC

15108 TGTACTCATGGTACCCTGGGGGCCCCGCTTCCCACCCCTGACAGGCAAAGACAGAAAGTC
TCTGGGAACACTGCCTGGTGGCCGCTGGGCATTTTCTTCTTTTTTCTTTTCTTTTT
AGAGATGGAATTTTGCTCTTGTCAACCCAGGCTTGAGTGCAATGGCGTTATCTTGGCTCAC
TGCAACCTCCACCTCTGGGGTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGTCGCTGAG
ATTACAGGTGCCACCACACCCAGCTAATTTTTGTATTTTTAGTAGATATTGGGTTTCACC
[A, C]
TGTGGCCAGGCTGGTGTCAAACCTCCTGACCTCAGGTGATCCACCTACCTTAGCCTTCCA
AAGTGCTGGGATTACAAGCCTGAGCCACTGCGCCAGCCTGGGCATTTTCTTCTTGAT
GAGGTGCTACCATCTCCAGGGAAGCCACTGAACCCCCAAGGCCCTTCTCCATTTTCTGG
CTAAGATAGGACATGGCCATGGACTTTTGAACAACCCAGAGGGGGAACAGCAGTGAATT
TCCTGGGGAACCCAGGCAGCCAGGGCTAGCAAGGCTGGGGTGGCCATGGCAGTAATCCT

16274 CTTCCAGACTGTCTCCCTGCAGGAGCTGCAGCAGGACTTTGAGAATGCGCCCCCACC
CCCCAACAAACACAGACCCCGGCTAACGGCACCAGCGTGTCTATATCACCTTCAGCCC
TGACAGCTCCTCACCTGCCCAGAGTGAGCCACCAGCCTCCGCTGAGGCCCGGCGAGCC
CAGTGACATGCTGGCCAGCGTCCCACCCCTTCGTACCTTCCACACCCCTCATCCTGGACAT
GAGTGGAGTCAGCTTCGTGGACTTGATGGGCATCAAGGCCCTGGCCAAGGTGAGGCCCTC
[-, G]
GGGACAGCAAGCACCACCCACTCCACCCCTCCGCTCTGCTCTCCACATTCCCTTTCTGT
GGAGCCCTCATTTTCAGGAAGCTGAGGGAGGAAGCTCACTGGGGAGACTAACAGCTCCTAG
GAATCCCTCCTTTCCCAGACGCCACCAGGTTGAGACATTCTCCACAGAGCAGGCCCAGA
CGGCCCATGACAATGAGTGGCGGGACAAGTCTACAGAGTTTCAGGCCCTGTGCTCCCA
ACACCCAGCAGTGGCCATCCCAAGTCCCTCTCAGCCATCAGGAACCCACCCAGGTTCT

17424 AACATGGTGAAACCCCGCTCTACTAAAAATACAAAAATTAGCCAGGTGTGGTGACGGGC
CCCTGTAGTCCCAGTACTCGGTAGGCTGAGGCAGAGAATTGCTTGAACCCAGGAGGCGG
AGGTTGCAGTGAGCCAAGATCGCGCCACTGCACTCCAGCCTGGGCAACAGAGTGAAACTC
CATCTCAAAGAAAAAAGAAAAATCTAGCCCCACAAGAAGGGGCCATGGTGACTTT
AAGTGCCCGCCACGTTGGCAAAAGTCCATTTCCGCTCCACTTCCCAGAGAAACCGTCAGC
[C, T]
AACACTCCAGGGAGAAGTGGTGTGCTTTGCTGCTATTTTGTCTTTGGCTGCTGGGCTCT
CAGGGTTGCTTATTTGTTGGCTTCCCTCTGAAGTACGTTTTGTGAATCACTTTTGAGA
CCCACTCAGAACATTCTTTCTTTGCTCCCTACCCCAACAACACTTCTAGCTGAGCT
CCACCTATGGGAAGATCGCGCTGAAGGTCTTCTTGGTGAACATCCATGGTAAGAGAAAGA
GGACATTTAGGGACTGAAAGACTGGCAAGGAGTGTGGGGTAGGAACAGGTTGGTGGGGTC

17627 AATATCTAGCCCCACAAGAAGGGGCCATGGTGACTTTAAGTGCCCGCCACGTTGGCAAAA
GTCCATTTCCGCTCCACTTCCCAGAGAAACCGTCAGCCAACACTCCAGGGAGAAGTGGTG
TGCTTTGCTGCTATTTTGTCTTTGGCTGCTGGGCTCTCAGGGTTGCTTATTTGTTTGGC
TTCCCTCTGAAGTACGTTTTGTGAATCACTTTTGAGACCCACTCAGAACATTCTTTCC
TTTTGCCCTCCCTACCCCAACAACACTTCTAGCTGAGCTCCACCTATGGGAAGATCGGCGT
[G, A]
AAGGTCTTCTTGGTGAACATCCATGGTAAGAGAAAGAGGACATTTAGGGACTGAAAGACT
GGCAAGGAGTGTGGGGTAGGAACAGGTTGGTGGGGTCTGAATAGTGAGGAGGTTGGAAC
GAGAGCACCAGCTATCCCCACAAGCTGCTGCCTGCTCATAAAAGCTTCAGGTACAAGT

FIGURE 3, page 16 of 20

CCAAAGAGACTGGTCAGATTGCATAAACATCCTAGGGGCCCTTAGTGACAGAGTGGGGGTG
AGGAGGTCATGGAGTTACAGAAGGACAGCTAGGATTCTAATCTACCCATAACTAATTTG

18427 GGGTGCATATACACAGCCTCAAGGACGTGGCCACAGGGCAGCAGACATTTACATGACTAG
CATGTACGCAAAAGTGCAGAGATGTGGGAGCAAGTGCACACAGACACACAGGAGAATGTGA
AGGGGCACATACACACACACCCAGCTCCCTGCACTGGGTGAGACCCCTCCAGCAGGGCT
GCAGTTCCCAAGCTCCGCATGGCCACGTTCGGGGAGAGAATCTGCAGTGGCAATGACCTG
CTATGATATGTTCTGGAGTTAGAAGCAGTGGATTCTCCCCAACCTCACTGGACACCCCT
[T, C]
AGGAAACCATCTCTAGGATTAAGAGTAATCCACACAACTTCCAATGCCACACATTGGAA
GTTGCTGGAAGGTCTGGGAAAACAAGAGGAAGGATGGGTCTTGGGGGATAGAACTGGC
AGCGGCTCTTCAAGGATGGCTTAGGCTTTTCCACTCGAATCACCACAAAGTACTGACTC
CCTAAATCAAAGTCTCTCTGCTCTGGGTGAACTTCAGCATCCTCAAGTTCATGT
TGCCCTCTGCCGTCCAGAACTGATATTGCACTGCCAATGCCATGGCCCTCAGATACAGCA

18813 AGAGGAAGGATGGGTCTTGGGGGATAGAACTGGCAGCGGCCCTCTCAAGGATGGCTTAG
GCTTTTCCACTCGAATCACCACAAAGTACTGACTCCCTAAATCAAAGTCTCTCTGCTG
TCTGGGTGAAACTTCAGCATCCTCAAGTTCATGTTGCCCTCTGCCGTCCAGAACTGATA
TTGCACTGCCAATGCCATGGCCCTCAGATACAGCAAGAGCTGGGACCTCAGGCCCTCTCC
ATCCCTGCTCTGGTCTCACTATCTTCCCCACCCCACTCCAATCCACAATGGCTGTTAT
[C, G]
TTTCTGAAGGTGATCTTTTCTCCTTCTAGCCAGGTGTACAATGACATTAGCCATGGAGG
CGCTTTTGGAGTGGGAGTCTAGAATGCAAGCACGTCTTCCAGCATACATGACGCAGT
CCTCTTTGCCAGGCAAAATGCTAGAGACGTGACCCAGGACACAACTTCCAAGGGGTAAG
GTTCTTGCACTGGGGAATCCTAGGCTCCAAGGCACTGAAATAGCAGGACCAAGAGGCAT
TATTAGAAAGAACACAGGAGAAGGTTTAAAGTTCCAATATCAAGTCTGCCATTTCAAGTTT

19035 GGACCTCAGGCCTCTCCCATCCCTGCTCTGGTCTCACTATCTTCCCCACCCCACTCCA
ATCCACAATGGCTGTTATCTTTCTGAAGGTGATCTTTTCTCCTTCTAGCCAGGTGTACA
ATGACATTAGCCATGGAGGCGTCTTTGAGGATGGGAGTCTAGAATGCAAGCACGTCTTTT
CCAGCATACATGACGCAGTCTCTTTGCCAGGCAATGCTAGAGACGTGACCCAGGAC
ACAACCTCCAAGGGGTAAGGTTCTTGACCTGGGGAATCCTAGGCTCCAAGGCACTGAAA
[T, C]
AGCAGGACCAAGAGGCATTATTAGAAAGAACACAGGAGAAGGTTTAAAGTTCCAATATCAA
GTCTGCCATTTCAAGTTTCTGAATCTGTTTCTTATCTATAGAATGAGCACCATCAACTA
ACATTACCTACCTCTCTGCATTTTCTTTTATTTTGTGTTTAGGGTTAAATGATAATTACA
TCTTTTGTGTCACTTGAAGGCTTTTGTGATTGTAAAAATCTTTATCAATATAAGTTT
TCTGGTTGCACAAACACCCAAAGCATAGTAGAGCAGGCCACTCTGCTGGCATCGTTCCC

19182 AGGATGGGAGTCTAGAATGCAAGCACGTCTTCCAGCATACATGACGCAGTCTCTTTG
CCCAGGCAAAATGCTAGAGACGTGACCCAGGACACAACTTCCAAGGGGTAAGGTTCTTGC
ACCTGGGGAATCCTAGGCTCCAAGGCACTGAAATAGCAGGACCAAGAGGCATTATTAGAA
AGAACACAGGAGAAGGTTTAAAGTTCCAATATCAAGTCTGCCATTTCAAGTTTCTGAATCT
GTTTCTTATCTATAGAATGAGCACCATCAACTAACATTACCTACCTCTCTGCATTTTTT
[T, C]
TTTATTTTGTGTTTAGGGTTAAATGATAATTACATCTTTTGTGTCACTTGAAAGCACTTTG
TGTATTGTAAAAATCTTTATCAATATAAGTTTTCTGGTTGCACAAACACCCAAAGCATA
GTAGAGCAGGCCACTCTGCTGGCATCGTTCCCTGCCTCCTCCTCATCTCTTTCTAAAGG
GGGCTTTTCGGGAAGGGAGGGGAGGGGAGTAAGCCTACCCATTTTAACTTACCGGAGCTTA
GAGATTTCAAGCTGGTGAGGGATAAAGAGATTGGGTCTGAGTTTGTCTCAGCTTTTTGA

19508 TAATTACATCTTTTGTGTCACTTGAAAGCACTTTGTGTATTGTAAAAATCTTTATCAAT
ATAAGTTTTCTGGTTGCACAAACACCCAAAGCATAGTAGAGCAGGCCACTCTGCTGGCA
TCGTTCCCTGCCTCCTCCTCATCTCTTTCTAAAGGGGGCTTTTCGGGAAGGGAGGGGAGGG
GAGTAAGCCTACCCATTTTAACTTACCGGAGCTTAGAGATTTCAAGGCTGGTGAGGGATAA
AGAGATTGGGTCTGAGTTTTGTCTCAGCTTTTTGACATTTAATTTACTAGCTCAGTAAGT
[-, G, C]
ATACAAATGGGATACAAATAACACCATCTAAACTCCAGAAGACTGGGGAGTCAGAAAAA
TCCTACCTCCTTGGGGTCCCTGCCAGATCCCAAGTCTCTAGCCCTCAGGGTCCCCT
CCCAGCTCAGCTCCTGCCCTTGGCCTCCCAAGACTCTTGTGTGCCCCAGCCCTGGGTAA
AAACCTCCCCTGCCCTCTGTGGGTGATAAGAAAGGCTTTTCTGGCCCTAGAGCAATGATT
TGCTCTTTGCCCTTAAAGAGACTGATGAAGGTGAAACCATCTGTTCTAAGTGTGAAAGACT

FIGURE 3, page 17 of 20

008221" 6855h460

19571 AGTTTTCTGGTTGCACAAACACCCAAAGCATAGTAGAGCAGGCCACTCTGCTGGCATCG
TTCCCTGCCTCCTCCTCATCTCTTTCTAAAGGGGGCTTTGCGGAAGGGAGGGGAGGGGAG
TAAGCCTACCCATTTTAACTTACCGAGCTTAGAGATTTAGGCTGGTGAGGGATAAAGA
GATTGGGTCTGAGTTTTGTCTCAGCTTTTGTACATTTAATTTACTAGCTCAGTAAGTCAT
ACAAATGGGATACAAATAACACCATCTAAACTCCAGAAGACTGGGGAGTCAGAAAAATC
[T, G, C]
TACCTCCTTGGGGTCCCTGCCAGATCCCCAGTCATCTCTAGCCCTCAGGGTCCCTCCC
AGCTCAGCTCCTGCCCTTGGCCTCCCAAGACTCTTGTGTGCCCCAGCCCTGGGTAAAAA
CCTCCCTGCCCTCTGTGGGTCTAAGAAAGGCTTTTCTGGCCCTAGAGCAATGATTTGC
TCTTTGCCCTTAAGAGACTGATGAAGGTGAAACCATCTGTTCTAAGTGCTGAAAGACTGCC
CAGGAACACACAGGGCGCTGGCTCCTGCCCTCCATGCCTAGAGGGAAACCTGGGGAAAC

20147 GCCTAGAGGGAAACCTGGGGAAACAACGGGCTTTCTGCTTCGTGAAATTTGTCCGCAG
AGCAAAGAGGGAGATTCTGGAGGAAGCTGCATTAGTTGTTAGTGCCCTAATCATGTTTCAG
CTACTCTAGTTGGTATGTATACTTGATTAGTCATAGCACTTATAAATAATTTATATTTTA
TATAATATATACTTACATATTATAGACCATTACAGATACAAATCACACACATAAACACA
CACCTTTTCAACAGCATTGTGAGGGACAAAGCAGGCAAAGTGAGGCTGGTTATCAGACTT
[T, G]
AACAGATTAGAAAATATATTCCCAGGAGGACAGGAATCCCCAAGGTCAGGCAGCTAGCC
AATAGTTTTTCTAAGCTGAGTAAAACCTTCCCTGCCTCTAACGGCCACAAAGAGGGGAA
GACCGCGATACACACCTGTCTGGTATAAGGGGGAAGACCACAGCCGTGCTGTTTTTGTGA
GGCAGGTAAGGGGAAGGGGCAAGAGGATAAGTCATGTGTGAGGAAGCAGCGTCCAACCAGA
GCCGGCCACCTGTCCCTTTTCTGCCACCATGCACCAACTTTGCTGTTCACTCACTGAAG

20180 TTCCTGCTTCGTGAAATTTGTCCGCAGAGCAAAGAGGGAGATTCTGGAGGAAGCTGCATT
AGTTGTTAGTGCCCTAATCATGTTTCAGCTACTCTAGTTGGTATGTATACTTGATTAGTCA
TAGCACTTATAAATAATTTATATTTTATATAATATATACTTACATATTATAGACCATTCA
CAGATACAAATCACACACATAAACACACACCTTTTCAACAGCATTGTGAGGGACAAAGCA
GGCAAAGTGAGGCTGGTTATCAGACTTTAACAGATTAGAAAATATATTCCCAGGAGGACA
[G, A]
GAATTCCTCCCAAGGTCAGGCAGCTAGCCAATAGTTTTTCTAAGCTGAGTAAAACCTTCCCT
GCCTCTAACGGCCACAAAGGAGGGGAAGACCGCGATACACACCTGTCTGGTATAAGGGGG
AAGACCACAGCCGTGCTGTTTTTGTGAGGCAGGTAAGGGGAAGGGGCAAGAGGATAAGTCA
TGTGTGAGGAAGCAGCGTCCAACCAGAGCCGGCCACCTGTCCCTTTTCTGCCACCATGC
ACCAACTTTGCTGTTCACTCACTGAAGCTCATTCTGCACTGGCTTCCTCCCTTCCAGGCT

20584 TGTCTGGTATAAGGGGGAAGACCACAGCCGTGCTGTTTTTGTGAGGCAGGTAAGGGAAGG
GGCAAGAGGATAAGTCATGTGTGAGGAAGCAGCGTCCAACCAGAGCCGGCCACCTGTCCC
TTTTCTGCCACCATGCACCAACTTTGCTGTTCACTCACTGAAGCTCATTCTGCACTGGC
TTCCTCCCTTCCAGGCTCCAGGGGATGCTGAGCTCTCCTTGTACGACTCAGAGGAGGACA
TTCGCAGCTACTGGGACTTAGAGCAGGTGAGCTGAGGGAAGGGGCTGTGAGGGTGGGAGC
[A, T]
GGGCGAAGAGGGGAAGGATGGGGTCTGCTGTCAAATACAAGGCGTTCACTCAGCTGTCTCA
CCTCCAGCCAGAGCAGTCACATTCAAGGCCACAAAGATTTGTGGTCATCTTTGTTTTTTT
TTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
AGACTGGAATGCAGTGGCATGATCTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCCAGA
GGTTCTCCTGCCTCAGCTCCCGAGTAGCTGGGACTTCAGGCTGCGCCAGCTAATTTT

20717 ATGCACCAACTTTGCTGTTCACTCACTGAAGCTCATTCTGCACTGGCTTCCTCCCTTCCA
GGCTCCAGGGGATGCTGAGCTCTCCTTGTACGACTCAGAGGAGGACATTTCGAGCTACTG
GGACTTAGAGCAGGTGAGCTGAGGGAAGGGGCTGTGAGGGTGGGAGCAGGGCGAAGAGGG
GAAGGATGGGGTCTGCTGTCAAATACAAGGCGTTCACTCAGCTGTCTCACCTCCAGCCAG
AGCAGTCACATTCAAGGCCACAAAGATTTGTGGTCATCTTTGTTTTTTTTCTTTTCTTTT
[T, C]
CTTTTTTTTTTTTTTTTTTAAATTTGAGACAAAGTCTCACTCTATCACCCAGACTGGAATGCA
GTGGCATGATCTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAGAGGTTCTCCTGCCT
CAGCCTCCCGAGTAGCTGGGACTTCAGGCCTGCGCCAGCTAATTTTTGTATTTTTTAGTA
GAGACAGCTTTTACCATGTTGGCTGGGCTGGTCTCGAACTTCGATCTCAAGCAATCTG
CCTGCCTCGGTCTCCTAAGTGCTGGATTACAGGCATAAGCCACGATGCCTGGCCTTTGT

20894 GGGGAAGGATGGGGTCTGCTGTCAAATACAAGGCGTTCACTCAGCTGTCTCACCTCCAGCC
CAGAGCAGTCACATTCAAGGCCACAAAGATTTGTGGTCATCTTTGTTTTTTTTCTTTTCT
TTTTCTTTTTTTTTTTTTTTTAAATTTGAGACAAAGTCTCACTCTATCACCCAGACTGGAA

FIGURE 3, page 18 of 20

GGCTGATCCAGATTGGGTCTTCCTGCACAGGAAGACTCTTTAACACCCTTAGGACCTCAG
GCCATCTTCTCCTATGAAGATGAAAATAGGGGTTAAGTTTCCATATGTACAAGGAGGTA
TTGAGAGGAACCCTACTGTTGACTTGAAAATAAATAGGTTCCATGTGTAAGTGTGTTTGT
AAATTTCACTGGAAATGCACAGAAAATCTTCTGGCCTCTCATCACTGCTTTTCTCAAGCT
TCTTCAGCTTAACAACCCCTTCCCTAACAGGTTGGGCTGGCCAGCCTAGGAAAACATCC

24210

TCACTGCTTTTCTCAAGCTTCTTCAGCTTAACAACCCCTTCCCTAACAGGTTGGGCTGGC
CCAGCCTAGGAAAACATCCCCATTTCTAACTTCAGCCAGACCTGCGTTGTGTGTCTGTGT
GTTGAGTGAGCTGGTCAGCTAACAAGTCTTCTTAGAGTTAAAGGAGGGGGTGCTGGCCAA
GAGCCAACACATTCTTGGCCAGGAGCATTGCTTTTCTGTGAATTCATTATGCCATCTGG
CTGCCAATGGAACTCAAACTTGAAGGCGAAGGACAATGTTATCTGGGATTCACCGTGC
[A, C]
CAGCACCCGAAGTGCCAAATTCAGGAGGACAAGAGCCTTAGCCAATGACAACTCACTCT
CCCCTACTCCACCTCCTTCCAAGTCCAGCTCAGGCCAGGAGGTGGGAGAAGGTCACAGA
GCCTCAGGAATTTCCAAGTCAGAGTCCCCTTTGAACCAAGTATCTAGATCCCCTGAGGAC
TTGATGAAGTGATCCTTAACCCCCAAGTAATCATTAAACCCCAAGACCGCTCAGAAGTG
AAGGAGATTGTTGACCCAGTGACCTGGAGTTGAGGCTCAGGGAGAGATCTGCCACATGTC

6356450